Implementation of a Neurophsiology-Based Coding Strategy for the Cochlear Implant

A report submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Master of Science

presented by
ONUR BABACAN

born 10 Februrary 1986
citizen of Turkey

accepted on the recommendation of
Prof. Dr. Norbert Dillier, Thesis supervisor
Dr. Wai-Kong Lai, Thesis co-supervisor
Prof. Dr. János Vörös, Track advisor

2010

Contents

Chapter 1

Abstract

Refractory State Coding (RSC) is a new coding strategy based on a functional model of the stimulated neural population. Our hypothesis is that RSC stimulation would convey the information contained in acoustic signals more effectively, improving sound perception and hearing performance for speech in noise and music. Two main factors that RSC takes into account are channel interaction [1] and refractory properties [2] of the stimulated neural population. They can be characterized by electrophysiological measurements of the evoked compound action potential (ECAP) using spread of excitation (SoE) and recovery function characterization paradigms respectively [3]. Using this information, for a given stimulus sequence, it is possible to calculate the refractory state of each stimulation site at any given time. In RSC, the stimulus is shaped according to the refractory states of stimulation sites. The spectral representation of the input sound is weighted by the refractory recovery information as well as the electric field distribution function before the next stimulus is selected. The Nucleus 24 and Nucleus Freedom family of cochlear implants incorporate Neural Response Telemetry (NRT) circuitry which is able to conveniently measure the ECAP from the implanted intracochlear electrodes, allowing the model to be custom-fitted to a patient. A software implementation of the standard ACE strategy for the Nucleus Cochlear Implant system is available in the Nucleus Matlab Toolbox. We implemented the RSC strategy in a compatible fashion in Matlab.

Chapter 2

Background

2.1 Cochlear Implant

2.1.1 Normal Hearing, Deafness

The human ear mainly consists of three parts: outer, middle and inner ear.

Sound undergoes a series of transformations as it travels through the outer ear, middle ear, inner ear, auditory nerve, and into the brain. The outer ear picks up acoustic pressure waves, which are converted to mechanical vibrations by a series of small bones in the middle ear. In the inner ear, the cochlea, a snail-shaped cavity filled with fluid, transforms the mechanical vibrations to vibrations in fluid. Pressure variations within the fluid of the cochlea lead to displacements of a flexible membrane, called the basilar membrane. These displacements contain information about the frequency of the acoustic signal. Attached to the basilar membrane are hair cells, which are bent according to the displacements of the basilar membrane. The bending of the hairs releases an electrochemical substance that causes neurons to fire, signaling the presence of excitation at a particular site in the inner ear. These neurons communicate with the central nervous system and transmit information about the acoustic signal to the brain [10].

The hair cells in conjunction with the basilar membrane are responsible for translating mechanical information into neural information. If the hair cells are damaged, the auditory system has no way of transforming acoustic pressure waves (sound) to neural impulses, resulting in hearing impairment. The hair cells can be damaged by certain diseases (e.g., meningitis, Meniere's disease), congenital disorders, by certain drug treatments, or by many other causes. Damaged hair cells can subsequently lead to degeneration of adjacent auditory neurons. If a large number of hair cells or auditory neurons are damaged, then the condition is called profound deafness. Research has shown that the most common cause of deafness is the loss of hair cells, rather than the loss of auditory neurons [10]. This is the basis of the cochlear implant approach, as neurons can be directly excited through electrical stimulation, bypassing the part of the human ear signal path beginning from outer ear and ending after the hair cells.

2.1.2 Cochlear Implant

A cochlear implant (CI) is a surgically implanted electronic device that provides a sense of sound to a person who is profoundly deaf or severely hard of hearing. Several cochlear implant devices have been developed over the years. All the implant devices have the following features in common: a microphone that picks up the sound, a signal processor

that converts the sound into electrical signals, a transmission system that transmits the electrical signals to the implanted electrodes, and an electrode or an electrode array that is implanted into the cochlea [10].

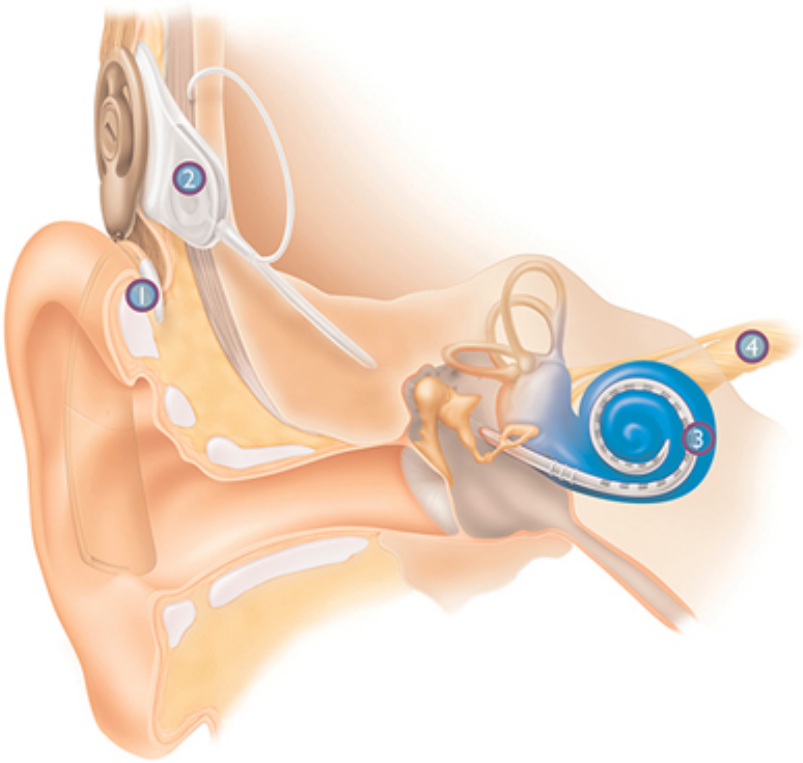


Figure 2.1: *Components of a cochlear implant system. The Nucleus Freedom system is illustrated, but all present-day implant systems share the same basic components. (1) The Speech processor captures, digitizes and encodes sound. (2) The speech processor transmits the encoded sound through the coil to the implant just under the skin. (3) The implant transmits the encoded sound along the electrode array which is positioned in the cochlea. (4)The electrodes stimulate the cochleas hearing nerve fibers, which relay the signals to the brain to produce hearing sensations.*

The implants employing an electrode-array exploit a property of the basilar membrane called tonotopic organization (also referred to as a "frequency-to-place" mapping). In this configuration, stimulation from each electrode corresponds to a localized band of the available human hearing spectrum. The speech processor processes the microphone output in order to create a stimulus pattern to be sent to the electrode array. The algorithm by which this processing is done is called a "speech coding strategy"

2.1.3 Cochlear Nucleus Freedom System

2.2 Speech Coding Strategies

2.2.1 Overview

A speech coding strategy is an algorithm that maps sound to an electrode stimulation pattern. Speech coding strategies play an extremely important role in maximizing the user's overall communicative potential, and different speech processing strategies have been developed over the past 30 years to mimic firing patterns inside the cochlea as naturally as possible [15]. Some of these techniques were aimed at preserving waveform information, others were aimed at preserving envelope information, and yet others were aimed at preserving spectral features (e.g., formants) [10]. We will discuss "N-of-M" strategies, with which we are primarily concerned within the context of our work, and take a brief look at other approaches.

2.2.2 N-of-M Strategies

N-of-M strategies separate speech signals into M subbands and derive envelope information from each band signal. N bands with the largest amplitude are then selected for stimulation (N out of M). The basic aim here is to increase the temporal resolution by neglecting the less significant spectral components and to concentrate on the more important features. These strategies have demonstrated either a significant improvement or at least user preference over conventional CIS-like(see §2.2.3) strategies [12].

We will examine ACE (Advanced Combinational Encoder) and MP3000 strategies, both of which are N-of-M type. ACE is the standard strategy used in Nucleus implant. MP3000 is a variation of ACE, and our RSC approach has similarities to MP3000, making it of interest to us.

ACE Strategy

The main mechanism of ACE is the mapping of the power spectrum of the signal to electrodes, where only N out of M ($N < M$) electrodes with the highest amplitude are activated. A block diagram of the process is presented in Figure 2.2.

The signal from the microphone is first pre-emphasized by a filter that amplifies the high-frequency components in particular. Adaptive-gain control (AGC) is then used to limit

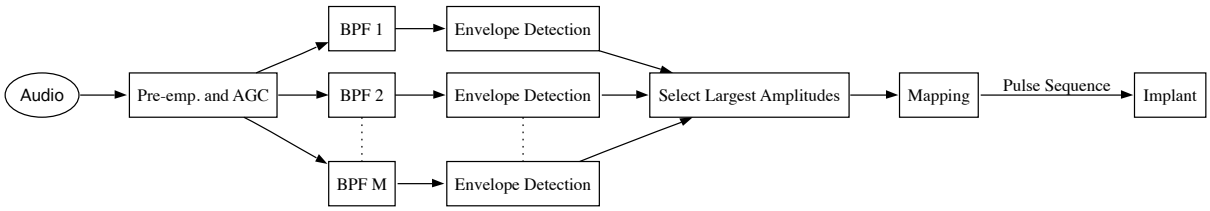


Figure 2.2: *Block Diagram of ACE Strategy*

distortion of loud sounds by reducing the amplification at the right time.

Afterwards, the signal is digitized and sent through a filter bank. ACE does not explicitly define a certain filter bank approach. The frequency bounds of the filter bank are linearly spaced below 1000 Hz, and logarithmically spaced above 1000 Hz.

An estimation of the envelope is calculated for each spectral band of the audio signal. The envelopes are obtained by computing the magnitude of the complex output. Each band pass filter is allocated to one electrode and represents one channel. For each frame of the audio signal, N electrodes are stimulated sequentially and one cycle of stimulation is completed. The number of cycles/second thus determines the rate of stimulation on a single channel, also known as channel stimulation rate [12].

The bandwidth of a cochlear implant is limited by the number of channels (electrodes) and the overall stimulation rate. The channel stimulation rate represents the temporal resolution of the implant, while the total number of electrodes M represents the frequency resolution. However, only N out of M electrodes ($N < M$) are stimulated in each cycle, therefore a subset of filter bank output samples with the largest amplitude is selected. If N is decreased, the spectral representation of the audio signal becomes poorer, but the channel stimulation rate can be increased, giving a better temporal representation of the audio signal. Conversely, if the channel stimulation rate is decreased, N can be increased, giving a better spectral representation of the audio signal.

Finally, the last stage of the process maps the amplitudes to the corresponding electrodes, compressing the acoustic amplitudes into the subjects dynamic range between measured threshold and maximum comfortable loudness level for electrical stimulation [12].

MP3000 Strategy

MP3000 (formerly known as PACE, Psychoacoustic ACE) is an ACE variant incorporating a psychoacoustic model [12]. The psychoacoustic model is used in selection of the N bands. A basic block diagram of the strategy is presented in Figure 2.3.

There are different classes of psychoacoustic models, the one used in MP3000 is a psychoacoustic-masking model. Such models describe masking effects that take place in a healthy auditory system. Psychoacoustic models have been successfully used within the field of audio coding in order to reduce bandwidth requirements by removing the less perceptually important components of audio signals. Because NofM speech coding strategies only select certain spectral elements of the audio signals, it can be speculated that a psychoacoustic model may ensure more effective selection of the most relevant bands than is achieved by merely selecting the spectral maxima, as with the ACE strategy. [12]

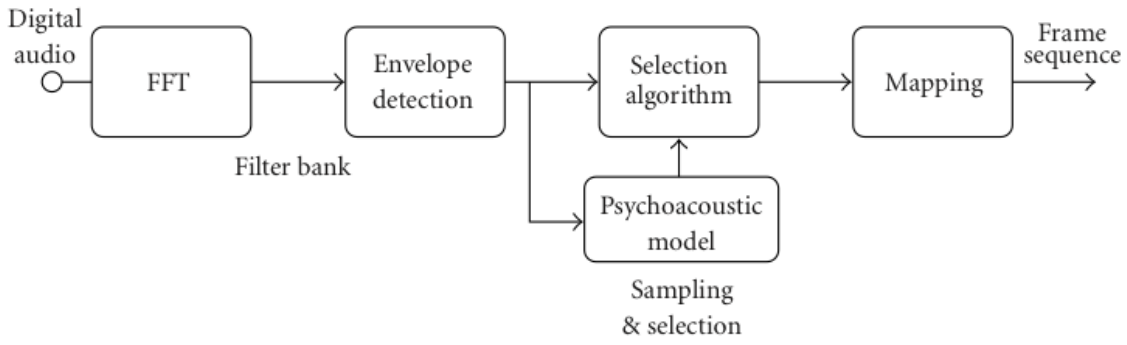


Figure 2.3: *Block diagram illustrating an "NofM" strategy incorporating a psychoacoustic model for selecting the N bands. The strategy may be termed the psychoacoustic ACE strategy [12].*

2.2.3 Other Approaches

Many other approaches to speech coding exist. A breakdown of the current coding strategies is given in Figure 2.4 [16]. The paradigms used in other strategies include pitch extraction, formant representation and rate modulation. We will not go into details of other strategies due to their irrelevance in the context of this work.

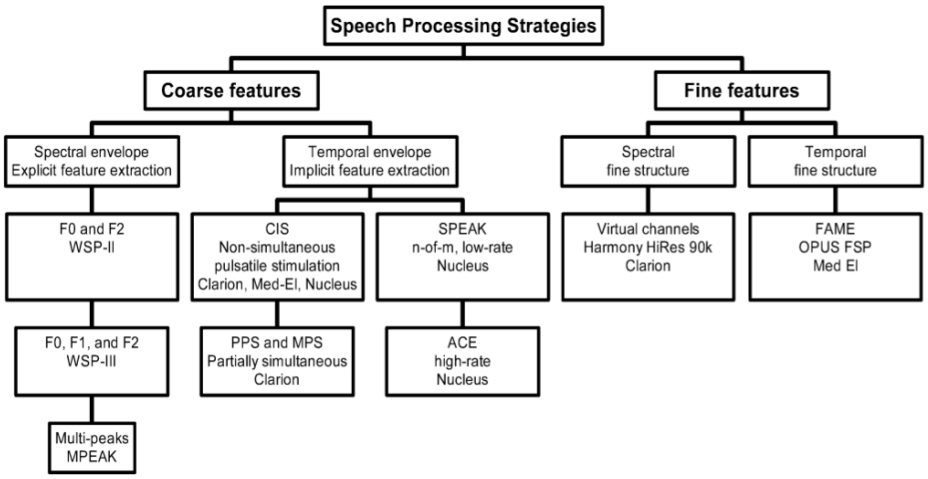


Figure 2.4: *Classifications of signal processing strategies in cochlear implants [16].*

Chapter 3

Refractory State Coding (RSC)

3.1 Introduction

With electric stimulation through a cochlear implant, the excited neural populations within the cochlea react in a much more deterministic manner than under acoustic stimulation. The stimulated neurons have a refractory behavior, which momentarily prevents the neural population from reacting to subsequent pulses that occur in quick succession. Furthermore, the spread of the stimulus electric field is much broader than the field of excitation with acoustic stimulation, the latter making use of the active tuning mechanisms of the basilar membrane. In standard coding strategies for the Cochlear Implant, neither the refractory properties nor the electric field spread are taken into consideration (an exception is MP3000, which recently has been evaluated in a large multicenter study involving 221 implanted subjects). The stimuli are generally presented on the selected electrodes as quickly as possible. This likely results in a pattern of activity in the neural population that is excessive, in both the time as well as space dimensions. The Refractory State Coding (RSC) strategy is intended to produce a less excessive and more natural activity pattern. As such RSC is based on a neurophysiological model taking into account both the refractory behavior and the spread of the stimulus electric field.

3.2 Background

As stated in §3.1, we employ a neurophysiological model. It is a spatiotemporal model, with the spatial component being represented by so-called "spread of excitation functions" and the temporal component being represented by "recovery functions". In this section we present these two foundations of our model.

3.2.1 Recovery Function

In the cochlear implant, when a single electrode is activated, a point-source-like electric field is generated. This electric field affects a population of neurons in the cochlea, evoking action potentials in them. This response of a population of neurons is called an "electrically evoked compound action potential" or "ECAP" for short.

In 1992, a bidirectional telemetry feature was implemented in the design of the Nucleus CI24M cochlear implant system that can be used to record electrically evoked compound action potentials (ECAPs). The action potentials that result from a stimulus applied on a given intracochlear electrode are recorded from a neighboring electrode, amplified, and then encoded for transmission via the radio frequency (RF) link back to the speech pro-

cessor. Custom (neural response telemetry (NRT) software, written at the Department of Otorhinolaryngology, University Hospital of Zürich, communicates with the speech processor to capture, process, store, and display the measurement data on a host personal computer. The NRT software controls the parameters of the stimulus used to evoke the response being measured, as well as the parameters used to perform the recording. [6]

The NRT software implements the masker-probe (also called forward masking) paradigm described by Brown et al [2] that is illustrated in Figure 3.1 and involves a sequence of 1) probe-only (A), 2) masker-followed-by-probe (B), and 3) masker-only (C) stimuli. The probe-only condition (A) yields the desired neural response plus an artifact from the probe. The masker-and-probe condition (B) with an appropriate masker advance yields stimulus artifacts from both masker and probe; the neural responses to the probe are absent or diminished because of forward masking. The masker-only condition (C) yields only the masker artifact. By using the subtraction method ($A - [B - C]$) to cancel the large masker stimulus artifacts found in each condition, one can extract the relatively small neural response [6].

NRT software allows for periodic ECAP measurements of a stimulation site at a given stimulus level, enabling modeling of the refractory behavior. The model obtained from such measurements is called a recovery function.

Former studies examined the recovery function using the forward masking technique. The stimulation paradigm is displayed in Figure 3.2 (left panel). The artifacts, which disturb the measurement of an unmasked electrical stimulus, the probe (A), are diminished subtracting a measurement of a masked stimulus (B). If the interstimulus interval, the masker-probe interval (MPI), is smaller than the absolute refractory period, the auditory nerve is not able to respond to the probe stimulus and the difference of both traces includes the neural response to the probe without a probe stimulus artifact. Two additional traces cancel the neural response and stimulus artifacts belonging to the masker stimulus (C) as well as the on and offset artifacts of the amplifier (D). Recovery functions were investigated by varying the MPI. An example is plotted in Figure 3.3. Such functions usually have a maximum at an MPI of about 500 μs and decrease for increasing MPIs. Former explorations of the recovery function suffered from this decrease since the peak-to-peak amplitude vanishes into the noise level for large MPIs. This complicates the reliable determination of a time constant. Furthermore, the existence of the maximum hampered the development of a model for the recovery function [11].

The recent suggestion of the modified forward masking technique by Miller et al. [2000] encourages to readdress this problem. They stated that the subtraction of neural responses to partially masked and unmasked stimuli used in standard NRT technique is inappropriate

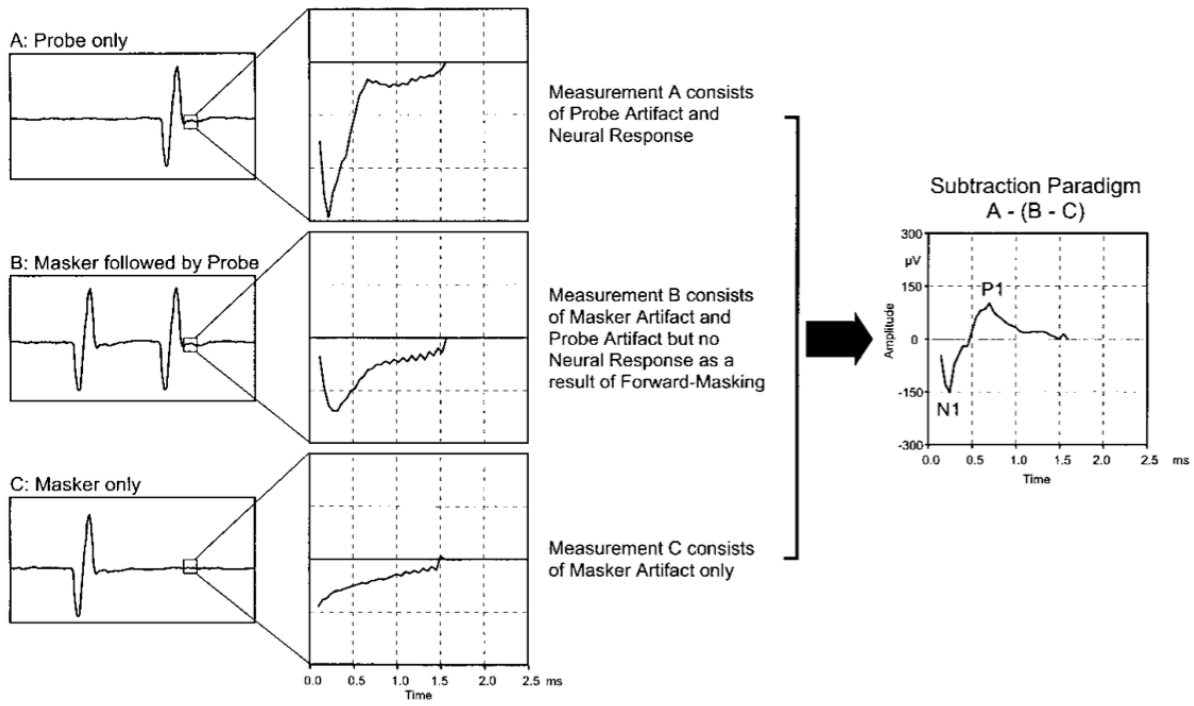


Figure 3.1: *Subtraction method for reducing masker stimulus artifact. Stimulation signals from buffers A, B, and C and their respective responses are shown on left, and subtraction paradigm* $(A - [B - C])$ *with resultant enhanced neural response is shown on right. First negative peak is labeled N1, and following positive peak is labeled P1. Amplitude of neural response is defined as N1-P1 peak- to-peak amplitude [6].*

due to different latencies of these neural responses [Finley et al., 1997; Miller et al., 2000]. Miller et al. [2000] bypass this problem by utilizing only masked stimuli. The stimulation pattern is shown in figure 1 (right panel). Two measurements of masked probe responses with different MPIs are subtracted (B1 - B2). If the shorter MPI, the reference MPI (refMPI), lies within the absolute refractory period, the corresponding recording (B2) includes no neural response to the probe stimulus but contains the stimulus artifact. In this case, the difference of both recordings does not include an artifact to the probe stimulus but the neural response to a masked probe with the MPI of trace B1. The stimulus artifacts and neural responses of the masker are diminished by the subtraction of sole masker stimulus recordings (C1 and C2). A recovery function can be derived varying the MPI. The reference MPI should be fixed and has to be taken smaller than the absolute refractory period. Charasse et al. [2003] varied the reference MPI between 125 and 500 μs within a single recovery function. Miller et al. [2000] tried 300 and 500 μs as a reference MPI. They decided to use that reference MPI which has the largest ECAP amplitude,

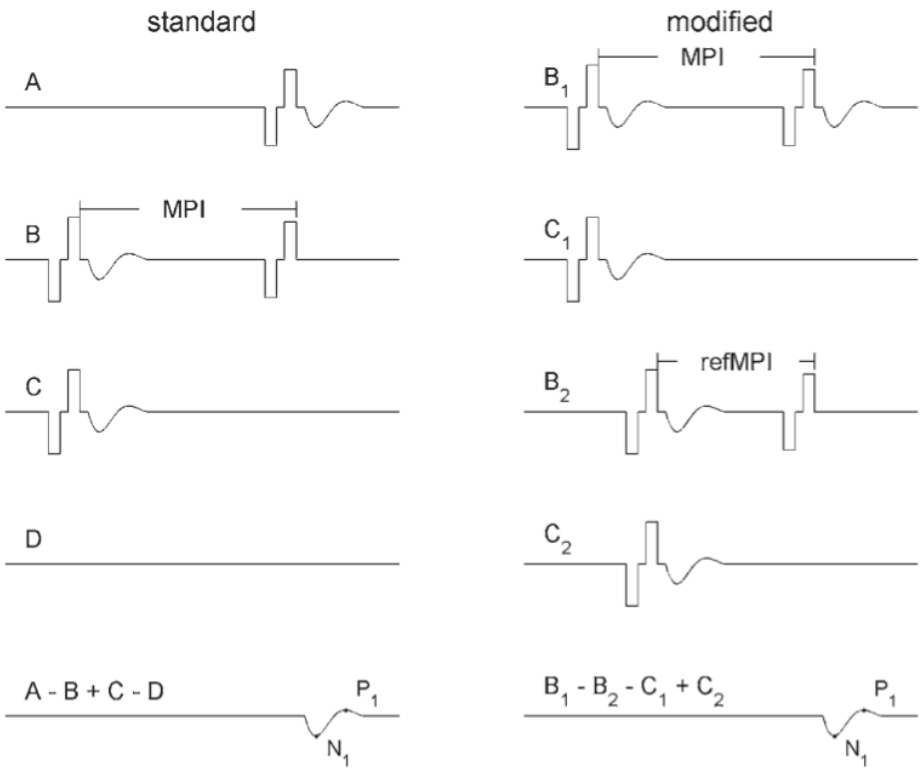


Figure 3.2: *Stimulation pattern of standard (left) and modified (right) forward masking technique. Idealized neural responses to the stimuli are drawn in the same trace, their amplitudes are enlarged to better visualize the output of the stimulation patterns. The ECAPs are plotted in the lowest line, their first negative and positive peaks are called N1 and P1, respectively. Adapted from Miller et al. [2000] [11].*

using this MPI in standard forward masking technique. According to them, the modified technique can be implemented using two measurements in standard forward masking with different MPIs. The measurements with masked stimuli (B) and sole masker stimulation (C) of the standard forward masking measurements of both MPI (1, 2) are subtracted from each other, written as ECAP = (B1 - C1) - (B2 - C2).

Measurements obtained using this approach can be used to fit an exponential model for data reduction, an example can be seen in Figure 3.4. The model is described by [11]:

$$\text{ECAP(MPI)} = \begin{cases} A.(1 - exp(-(\text{MPI} - T_0)/\tau)) & \text{for MPI} \geq T_0, \\ 0 & \text{for MPI} \leq T_0 \end{cases} \tag{3.1}$$

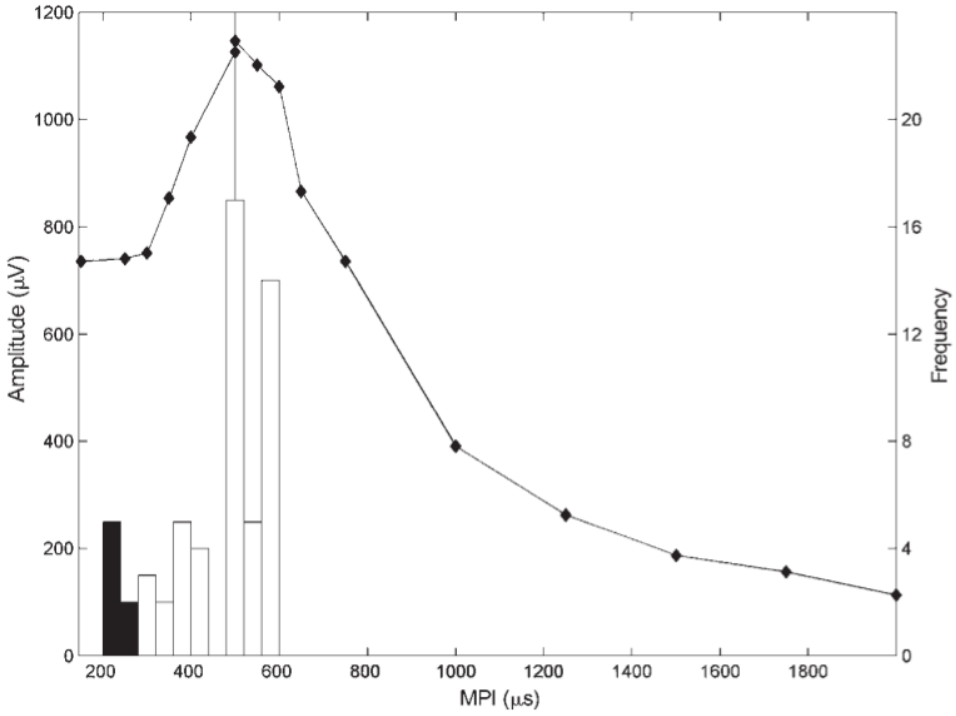


Figure 3.3: *Recovery function obtained at one stimulation site by the standard forward masking technique (black diamonds). Such recovery functions exhibit typically a peak at an MPI around 500 μs and decrease at increasing MPI. The histogram of 50 of these maxima at 71 stimulus sites (white bars) had a median of 500 μs (vertical line). Only distinctly detectable maxima are represented. Maxima were not detected in 21 recovery functions as they have similar ECAP amplitudes for small MPIs and therefore no unique maximum. The histogram of detected distorted wave- forms is displayed as well (black bars) [11].*

An example fitting is displayed in Figure 3.4. In RSC, we use this paradigm to model the refractory behavior.

3.2.2 Spread of Excitation Function

One likely determinant of performance with a cochlear implant is the degree of interaction that occurs when overlapping subsets of nerve fibers are stimulated by various electrodes of a multielectrode array. The electrically evoked compound action potential (ECAP) (see §3.2.1) can be used to assess physiological channel interaction. In [1], Abbas et. al. describe results from two different methods of analysis of ECAP channel interaction measures made by the Nucleus neural response telemetry (NRT) system. Using a forward-masking stimulus paradigm (see §3.2.1), masker and probe pulses are delivered through

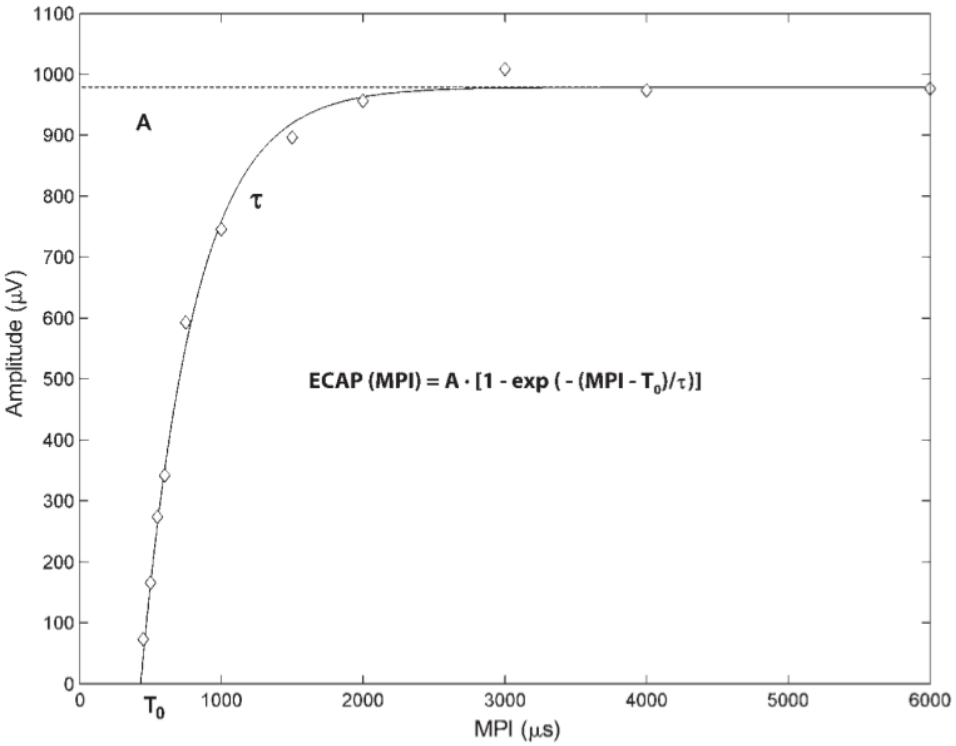


Figure 3.4: *Fitting the exponential model to an ECAP amplitude recovery function in the modified forward masking technique at a reference MPI of 300 μs. The neural response becomes measurable above the threshold T_0 which is an estimate of the absolute refractory period. The ECAP amplitude increases with a time constant τ up to an asymptotical limit, the saturation level A. (The fitted parameters are τ = 380.5 μs, T_0 = 430.6 μs, A = 978.8 μV) [11].*

different electrodes. The response to the probe is then dependent on the extent of overlap in the stimulated neural populations. The amplitude of response to the probe as a function of masker electrode position then reflects the degree of overlap between the population of neurons responding to the masker and those stimulated by the probe.

The lack of across-fiber independence in excitation is referred to as channel or spatial interaction and may impose significant limitations on performance in present cochlear implant designs [7] [14]. If each electrode stimulated the same neurons, then there would be no inherent benefit to multiple versus single-electrode implants. At the other extreme, if there were no overlap among stimulated neural channels, then one may assume that information provided by each stimulus channel is effectively transmitted to the central nervous system. The degree of spread of neural excitation and, consequently, the overlap of stimulated neurons from different electrodes may result in interactions that may

compromise the responses in a way that diminishes the information provided on the individual channels of stimulation. As a result, measures of the degree of overlapping response areas may be important in determining the performance of individual implant users and devices. The degree of spread of neural excitation in response to stimulation of a specific electrode or combination of electrodes may depend on electrode position relative to the stimulable neurons, the orientation of the electrodes and the resulting electric field, as well as the degree and pattern of neural survival. For instance, if all surviving neurons were remote to 2 monopolar stimulating electrodes, then significant overlap in the stimulated populations would be expected. Alternatively, if the mode of stimulation provides narrow current spread and stimulable neurons are close to the stimulating electrodes, one may expect relatively little overlap in the stimulated neural population, especially at low stimulus levels. [1].

A sample spread of excitation function from an actual measurement is presented in Figure 3.5

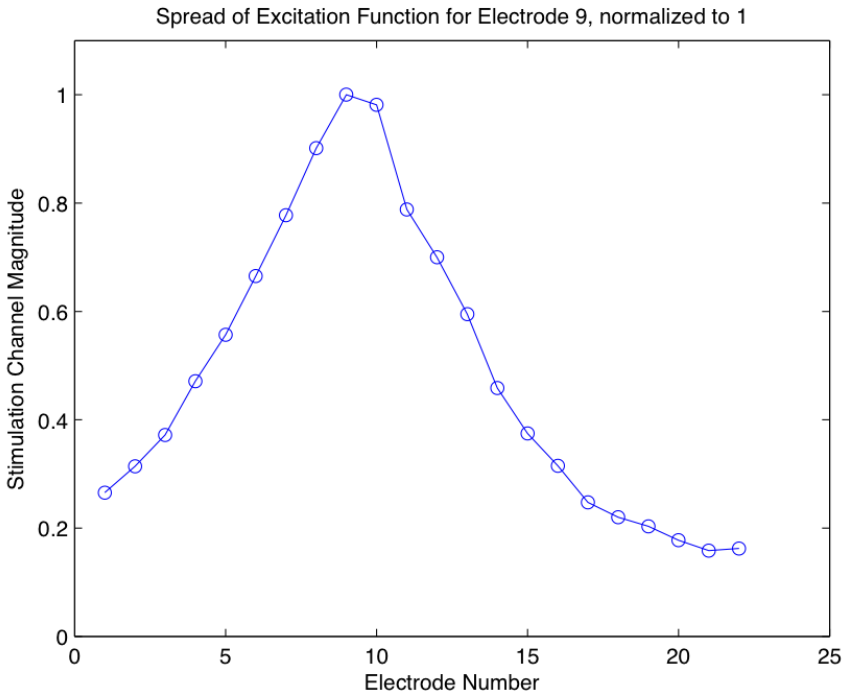


Figure 3.5: *Sample Spread of Excitation Function*

3.3 Algorithm

RSC strategy is a variation of the ACE strategy. It is similar to MP3000 in that it is a modification of the stimulus selection based on a model (see §2.2.2). Since RSC is a strategy in research state, it is based on the research ACE strategy available in the form of Nucleus Matlab Toolbox from Cochlear Corporation [3]. In this section we will examine this implementation of RSC strategy in detail. A block diagram is presented in Figure 3.7(b).

3.3.1 Research ACE Strategy

The research ACE strategy is designed to process signals that are already digitized. For this reason, the pre-emphasis filter and adaptive-gain controls (AGC) normally incorporated at the analog stage are not included in this setup [12]. A digital signal sampled at 16 kHz is sent through a filter bank without either pre-amplification or adaptive-gain control. The filter bank is implemented with an FFT (fast Fourier transform). The block update rate of the FFT is adapted to the rate of stimulation on a channel (i.e., the total implant rate divided by the number of bands selected N). The FFT is performed on input blocks of 128 samples (L = 128) of the previously windowed audio signal. The window used is a 128-point Hann window

$$w(j) = 0.5\left(1.0 - cos\left(\frac{2\pi j}{L}\right)\right), j = 0, \ldots, L-1. \tag{3.2}$$

The linearly-spaced FFT bins are then combined by summing the powers to provide the required number of frequency bands M, thus obtaining the envelope in each spectral band $a(z)(z = 1, ..., M)$. The real part of the jth FFT bin is denoted with $x(j)$, and the imaginary part $y(j)$. The power of the bin is

$$r^2 = x^2(j) + y^2(j), j = 0, \ldots, L-1. \tag{3.3}$$

The power of the envelope of a filter band z is calculated as a weighted sum of the FFT bin powers

$$a^2(z) = \sum_{j=0}^{L/2} g_z(j) r^2(j), z = 1, \ldots, M, \tag{3.4}$$

where $g_z(j)$ are set to the gains g_z for a specific number of bins and otherwise zero. This

mapping is specified by the number of bins, selected in ascending order starting at bin 2, and by the gains g_z as presented in Figure 3.6 [4] [3].

Band number z	1	2	3	4	5	6	7	8	9	10	11
Number of bins	1	1	1	1	1	1	1	1	1	2	2
Center freqs. (Hz)	250	375	500	625	750	875	1000	1125	1250	1437	1687
Gains g_z	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.68	0.68

Band number z	12	13	14	15	16	17	18	19	20	21	22
Number of bins	2	2	3	3	4	4	5	5	6	7	8
Center freqs. (Hz)	1937	2187	2500	2875	3312	3812	4375	5000	5687	6500	7437
Gains g_z	0.68	0.68	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65

Figure 3.6: *Number of FFT bins, center frequencies, and gains per filter band for $M = 22$.*

The envelope of the filter band z is

$$a(z) = \sqrt{\sum_{j=0}^{L/2} g_z(j) r^2(j)}, z = 1, \ldots, M. \tag{3.5}$$

In the sampling and selection block, a subset of $N (N < M)$ filter bank envelopes $a(z_i)$ with the largest amplitude are selected for stimulation.

The mapping block, determines the current level from the envelope magnitude and the channel characteristics. This is done by using the loudness growth function (LGF) which is a logarithmically-shaped function that maps the acoustic envelope amplitude $a(z_i)$ to an electrical magnitude

$$p(z_i) = \begin{cases} \frac{log(1+\rho((a(z_i)-s)/(m-s)))}{log(1+\rho)}, & s \leq a(z_i) \leq m, \\ 0, & a(z_i) < s, \\ 1, & a(z_i) \geq m. \end{cases} \tag{3.6}$$

The magnitude $p(z_i)$ is a fraction in the range 0 to 1 that represents the proportion of the output range (from the threshold T to the comfort level C). A description of the process by which the audio signal is converted into electrical stimuli is given in [13]. An input at the base-level s is mapped to an output at threshold level, and no output is produced for an input of lower amplitude. The parameter m is the input level at which the output saturates; inputs at this level or above result in stimuli at comfort level. If there are less than N envelopes above base level, they are mapped to the threshold level. The parameter ρ controls the steepness of the LGF, the selection of a suitable value for ρ is described in [4].

Finally, the channels z_i, are stimulated sequentially with a stimulation order from high-to-low frequencies (base-to-apex) with levels [12]:

$$l_i = T + (C - T)p_i. \tag{3.7}$$

3.3.2 RSC Physiological Model

As mentioned in previous sections, recovery and spread of excitation functions constitute the basis of our physiological model. In this section, we present the model's details.

We propose a model with separate states for each stimulation site. Each state has a value between 0 and 1 representing the fraction of neurons ready to fire, i.e. 0 means that all of the neurons at the site are in the refractory period (cannot fire), and 1 means that the site is completely available to receive stimulus.

Each state also has an associated recovery and spread of excitation function that are independent from other states, allowing for site-by-site modeling. The associated functions are in the range $[0, 1]$, representing the refractory behavior with the same paradigm as the state values.

The associated functions of each site can be determined by theoretical or statistical approaches, as well as be obtained by model fittings on NRT measurements specific to the implant user.

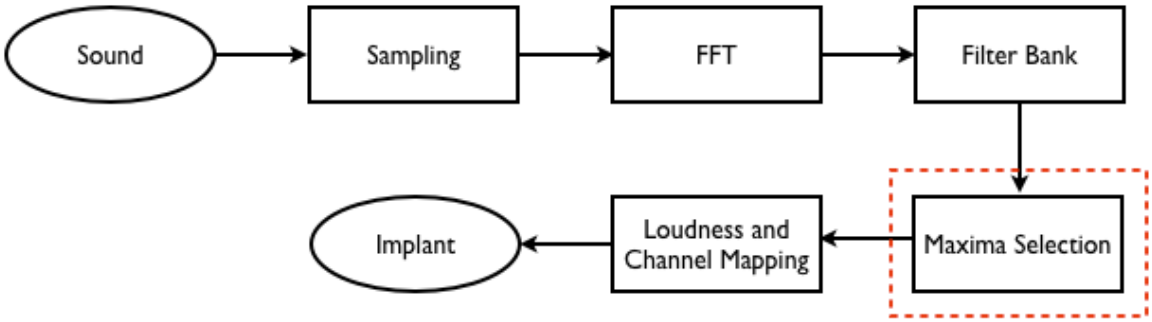
3.3.3 RSC Algorithm

Since RSC only replaces the "maxima selection" stage of the research ACE strategy, we focus on that stage of the strategy in this section (see Figure 3.7).

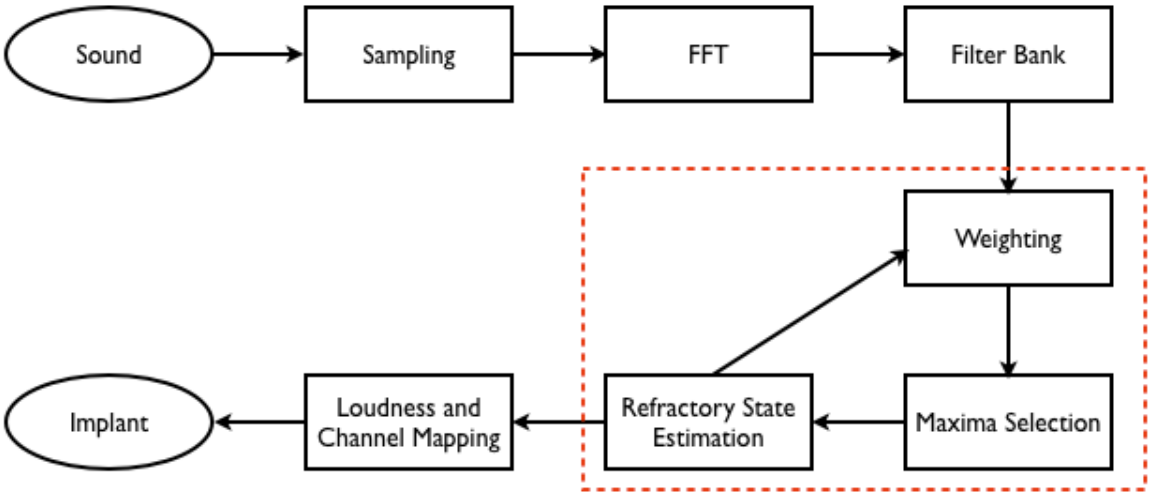
We first define the nomenclature that will be used to present the algorithm (see Table 3.1). All the vectors in Table 3.1 except where explicitly indicated in the following paragraphs are M-dimensional, and arranged in such a way that an index i refers to a value associated with stimulation site i in all vectors. In variable notations, we use superscripts to denote the iteration steps, subscrpts to denote the -th element of the vector and bold typeface to denote vectors. We use the dot notation for derivatives ($\mathrm{d}\, f(x)/\,\mathrm{d}\, x = \dot{f}(x)$) and operator "$\odot$" for pointwise product.

M is the number of stimulation sites that are available for stimulation. (By default $M = 22$ in the Nucleus Freedom system. In practice, usually $M < 22$ for clinical reasons.)

N is the number of stimulation pulses generated for each frame. In N-of-M strategy



(a) ACE Strategy



(b) RSC Strategy

Figure 3.7: *Comparison of ACE and RSC Strategy Block Diagrams*

literature, it is common to refer to this value as the "number of maxima", to which we will conform.

The excitability array, $\mathbf{e}$ is the collection of state values e_i corresponding to the refractory state of stimulation site i. It is used both as a state vector and a weighting vector.

The current frame vector, $\mathbf{f}$ with elements f_i, is the envelope estimate from which N stimulus pulses will be generated.

The weighted frame vector, $\mathbf{w}$, is the pointwise product of $\mathbf{e}$ and $\mathbf{f}$.

$\mathbf{p}$ describes the selected stimulus. It is a mathematical pair with elements (l, m). l and m correspond to the index of the stimulation site and the fraction of the site intended for

stimulation (i.e. stimulation magnitude), respectively.

$r_j(\hat{n})$ are real-valued functions in the range $[0, 1]$, with time as the independent variable (see 3.1), representing the recovery function belonging to stimulation site j.

$s_j(\hat{j})$ are real-valued functions in the range $[0, 1]$, with the set of stimulation sites as the independent variable, representing the spread of excitation function belonging to each stimulation site.

T and **C** are vectors with values in $[0, 255]$, representing the threshold and comfort level values per stimulation site. These values are obtained from the implant user's map.

Table 3.1: *List of variables*

M :	Number of available stimulation sites	n :	Iteration step index
N :	Number of selected maxima	j :	Stimulation site index
$\mathbf{e}$:	Excitability vector	$r_j(\hat{n})$:	Recovery function of site j
$\mathbf{f}$:	Current frame vector	$\mathbf{R}$:	Recovery vector
$\mathbf{w}$:	Weighted frame vector	$s_j(\hat{j})$:	Spread of excitation function of site j
$\mathbf{p}$:	Pulse	$\mathbf{S}$:	Stimulation vector
$\mathbf{T}$:	Threshold level vector	$\mathbf{C}$:	Comfort level vector

Before starting processing ($n = 1$) the model is initialized by setting

$$\mathbf{e}_i^n = 1, \quad i = 1, \ldots, M. \tag{3.8}$$

It is assumed here that the neuron population is completely polarized, i.e. fully recovered. This is a reasonable assumption because only insignificant depolarization is expected without the activation of the implant (a supporting point is that the justification of the implant's existence is dependent on this situation).

The weighted frame vector is then obtained by a point-wise multiplication of $\mathbf{e}$ and $\mathbf{f}$ such that

$$\mathbf{w}_i^n = e_i^n . f_i^n, \quad i = 1, \ldots, M \tag{3.9}$$

or, equivalently,

$$\mathbf{w}^n = \mathbf{e}^n \odot \mathbf{f}^n. \tag{3.10}$$

Based on this, the stimulus selection is done

$$\mathbf{p}_n = (l^n, m^n) \tag{3.11}$$

where

$$l^n = \quad argmax(\mathbf{w}^n) \tag{3.12}$$

$$m^n = \qquad w_{l^n}^n \tag{3.13}$$

The next stage of the algorithm keeps track of the refractory behavior caused by the selected stimulus. First the spread of the electric field is considered

$$\hat{s} = m^n s_{l^n}(\hat{j})(C_{l^n} - T_{l^n}) + T_{l^n} \tag{3.14}$$

where $\hat{s}$ describes what equivalent pulse level is delivered to each site by the spread of excitation. It is normalized per site (making it compatible with the excitability vector values) to obtain the stimulation vector

$$\mathbf{S}^n = \frac{\hat{s} - \mathbf{T}}{\mathbf{C} - \mathbf{T}} \tag{3.15}$$

Values of stimulation vector $\mathbf{S}$ represent what fraction of each site is being stimulated. Consequently, the stimulation vector will be used as a set of weights for corresponding recovery functions to obtain the recovery vector. Instead of directly using the recovery functions $\mathbf{r}(\hat{n})$, a vector consisting of their derivatives is defined

$$\dot{\mathbf{r}} \triangleq [\dot{r}_1(\hat{n}), \ldots, \dot{r}_M(\hat{n})]^T \tag{3.16}$$

and the recovery vector is updated

$$\mathbf{R}^n(\hat{n}) = \mathbf{R}^{\hat{n}-1}(n) + \mathbf{S}^n \odot \mathbf{r}^n(\hat{n} - n) \tag{3.17}$$

where

$$\mathbf{R}^0(\hat{n}) = \mathbf{0}_{M\times 1} \tag{3.18}$$

and this condition depends on the same assumption as (3.9).

Finally, the excitability array is updated:

$$\mathbf{w}^{n+1} = \mathbf{w}^n + \int_n^{n+1} \mathbf{R}^n(\hat{n})\,\mathrm{d}\,\hat{n}. \tag{3.19}$$

After the update, a new stimulus is selected and the iteration continues until N stimuli are selected for the current frame $\mathbf{f}$. It is then discarded and the next frame takes it place. During this change, it is critical that $\mathbf{R}$ and $\mathbf{w}$ are unmodified, as they contain the complete refractory state information.

To better illustrate how RSC algorithm works, we will present a sample iteration step-by-step. All values of the excitability vector are initialized with 1 ($\mathbf{e}^1 = \mathbf{1}_{M\times 1}$, we have the first envelope estimate (i.e. frame vector) $\mathbf{f}$ and consequently the weighted frame vector $\mathbf{w}^1$ equals $\mathbf{f}$ (see Figure 3.8). From $\mathbf{w}^1$, we choose the bin with the highest magnitude.

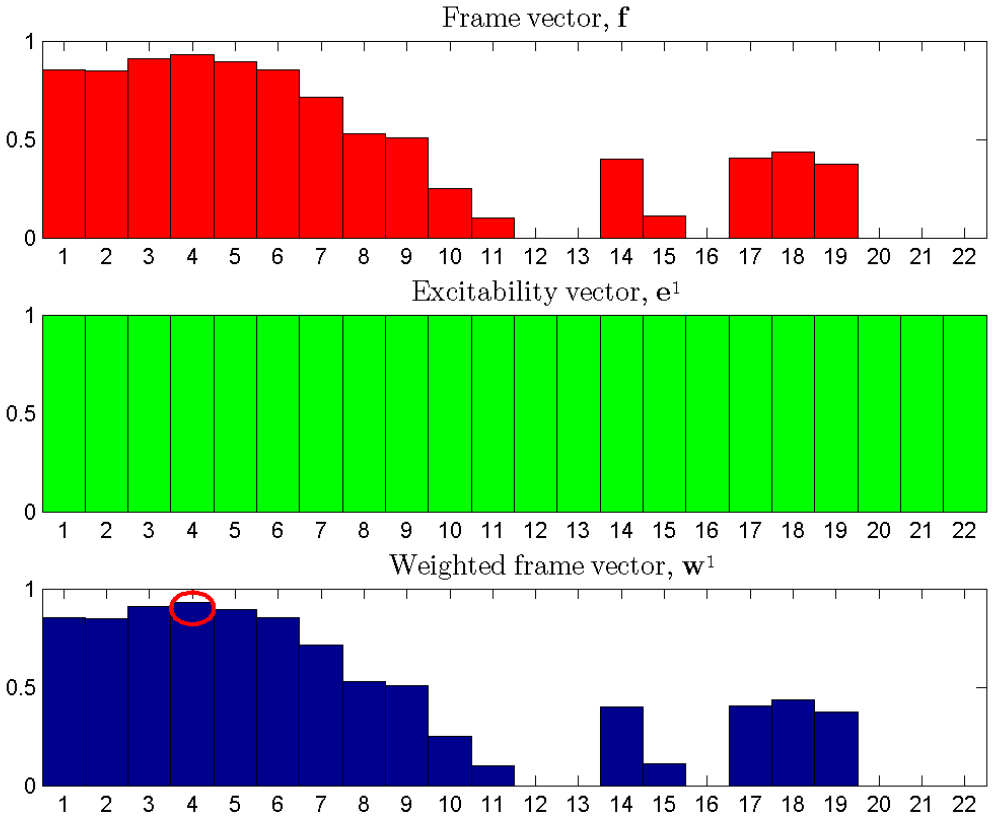


Figure 3.8: *Initial State,*$n = 1$, $\mathbf{e}^1, \mathbf{w}^1, \mathbf{f}$

In Figure 3.8 and consequent figures in this section, the red circle denotes the chosen bin

from $\mathbf{w}^n$ and the red arrows denote the bin selected in the previous iteration, with the arrow originating from the previous magnitude and the arrowhead ending at the current magnitude.

Inspecting the sequence of figures, it is observed that once a stimulus is sent, the excitability of its neighborhood drops, and slowly recovers over time due to recovery functions.

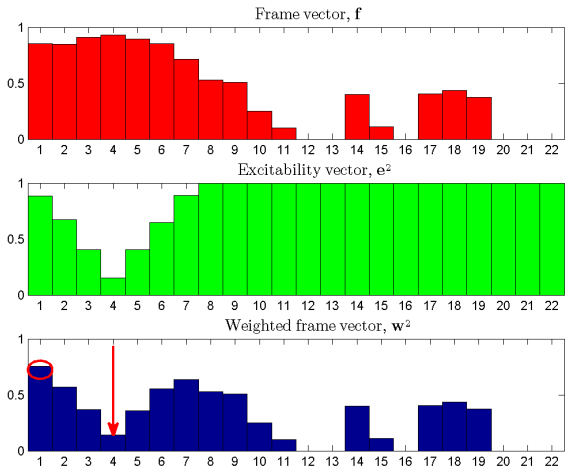


Figure 3.9: $n = 2$

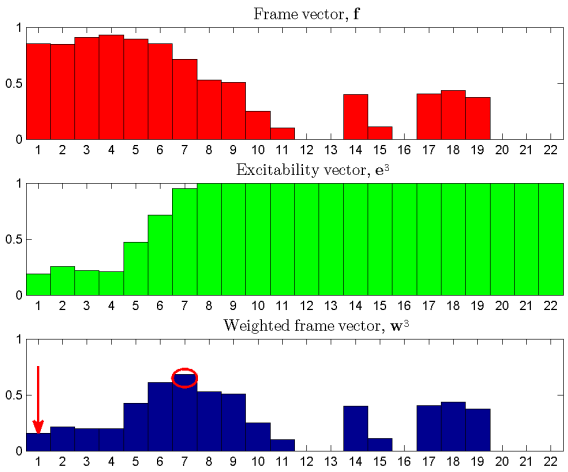


Figure 3.10: $n = 3$

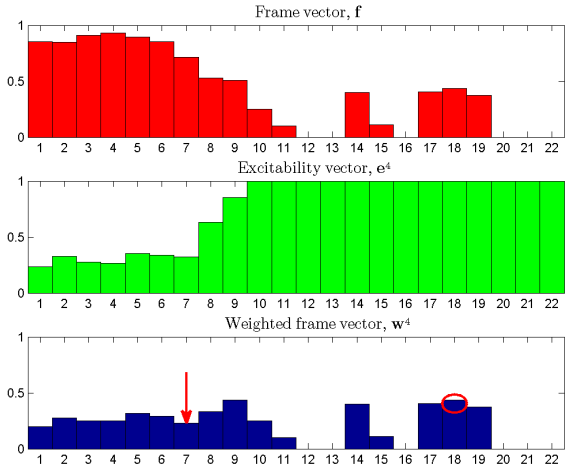


Figure 3.11: $n = 4$

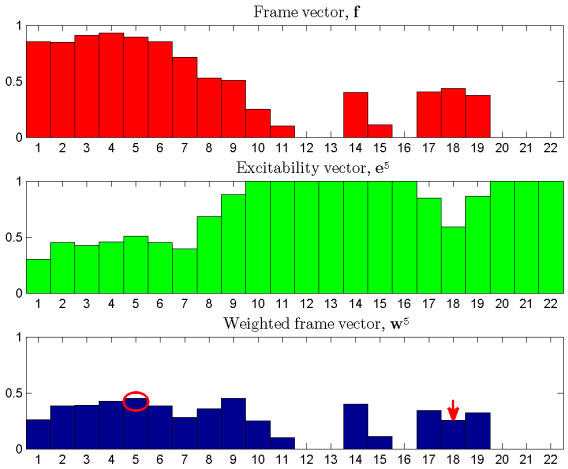


Figure 3.12: $n = 5$

In Figure 3.20, a new frame arrives since $N = 12$ maxima have already been chosen from the previous frame. Note that even though $\mathbf{f}$ changes, $\mathbf{e}$ is not reset, and the change in it is only due to the recovery functions as in any other iteration.

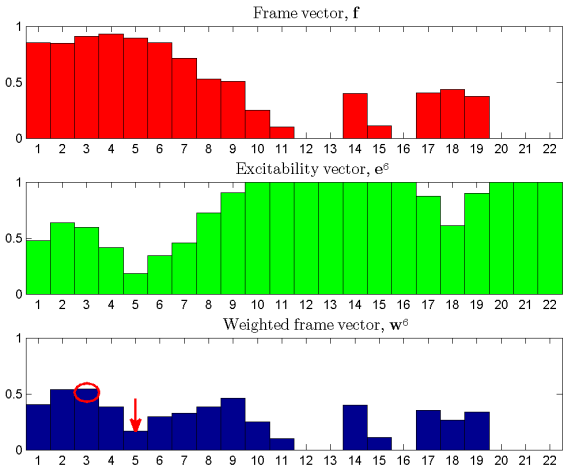


Figure 3.13: $n = 6$

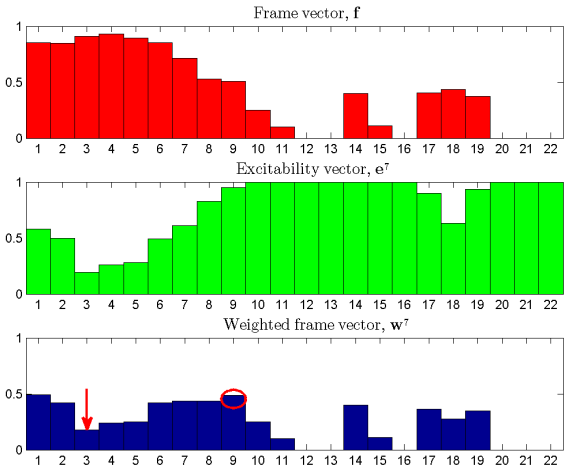


Figure 3.14: $n = 7$

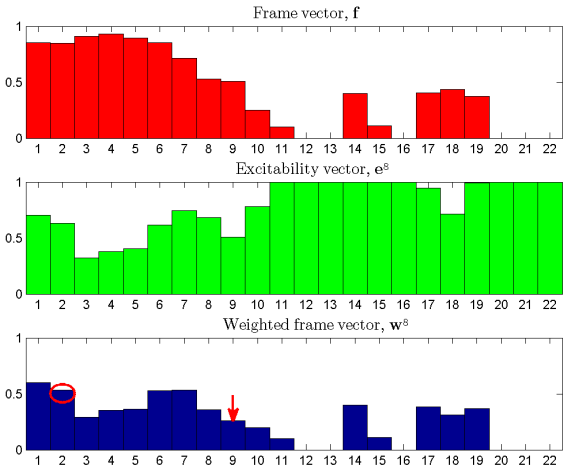


Figure 3.15: $n = 8$

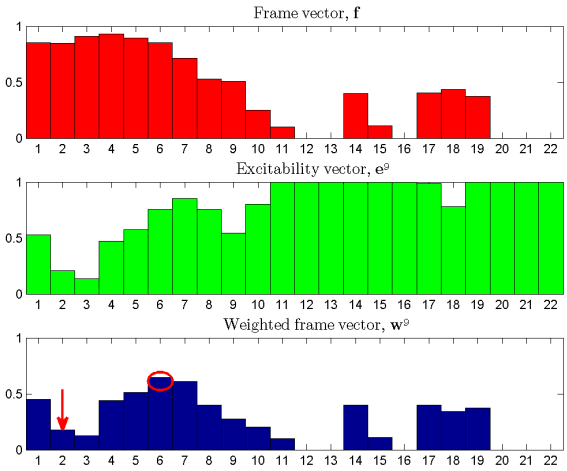


Figure 3.16: $n = 9$

To further demonstrate the different stimulus selection process of RSC in comparison to ACE, we present figures 3.21, 3.22, 3.23 and 3.24. These figures show the selected stimulus magnitude and channel for different number of selected maxima per frame and different strategies. For 12 maxima, we observe that RSC is the same with ACE in terms of stimulation site selection, only the magnitudes are different. As we decrease the number of maxima in the subsequent figures, we observe that ACE is limited to lower frequencies while RSC better represents the full spectral characteristics of the input frame.

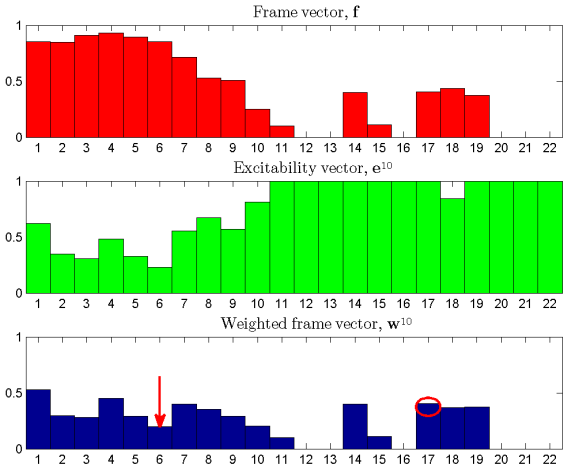


Figure 3.17: $n = 10$

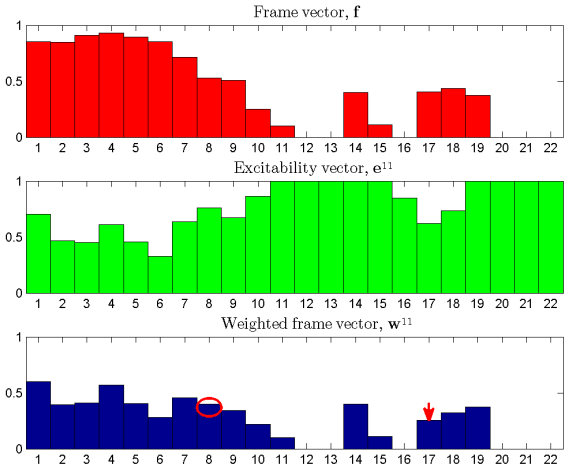


Figure 3.18: $n = 11$

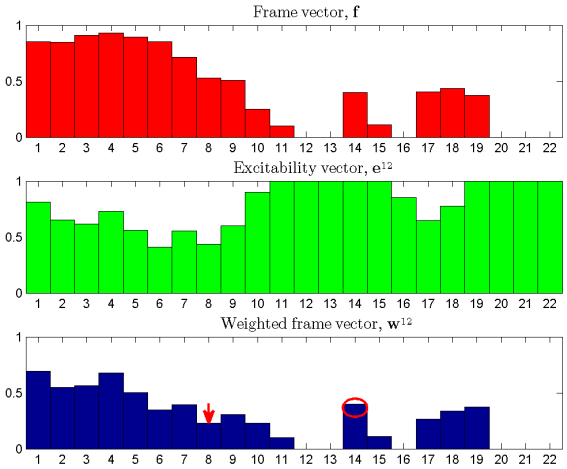


Figure 3.19: $n = 12$

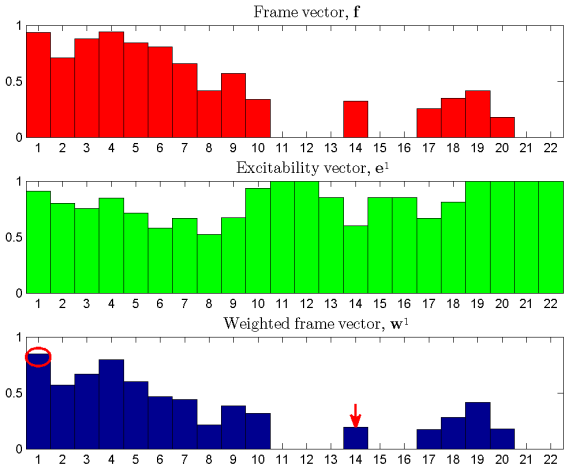


Figure 3.20: $n = 13$

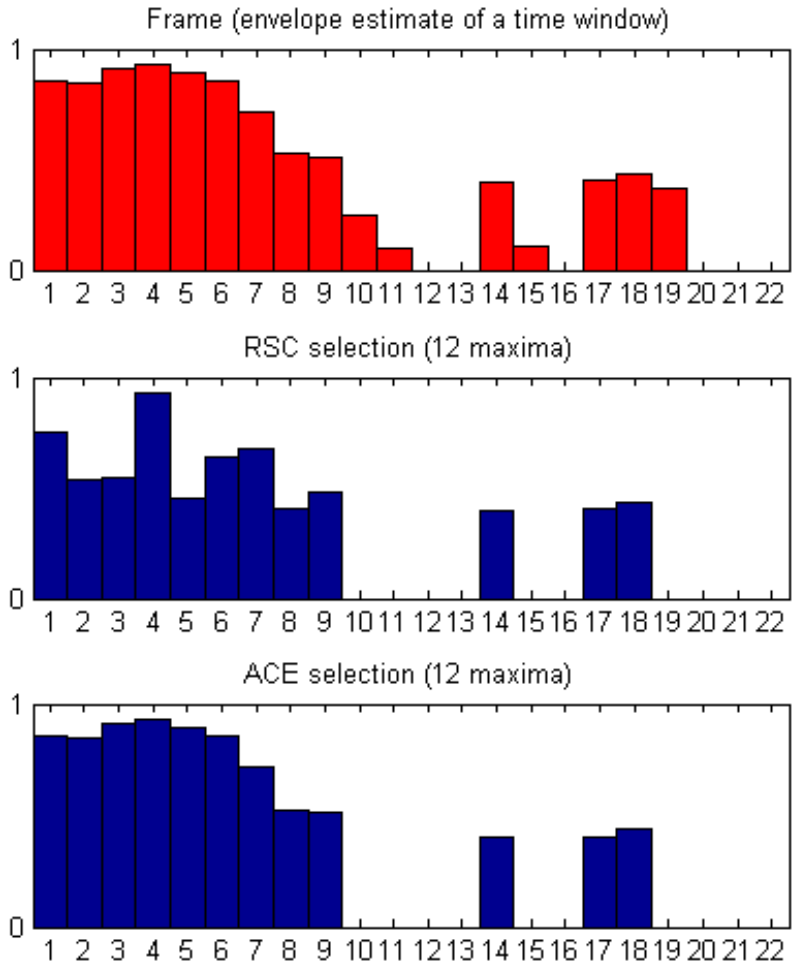


Figure 3.21: *Stimulation selection for ACE and RSC on the same frame, 12 maxima*

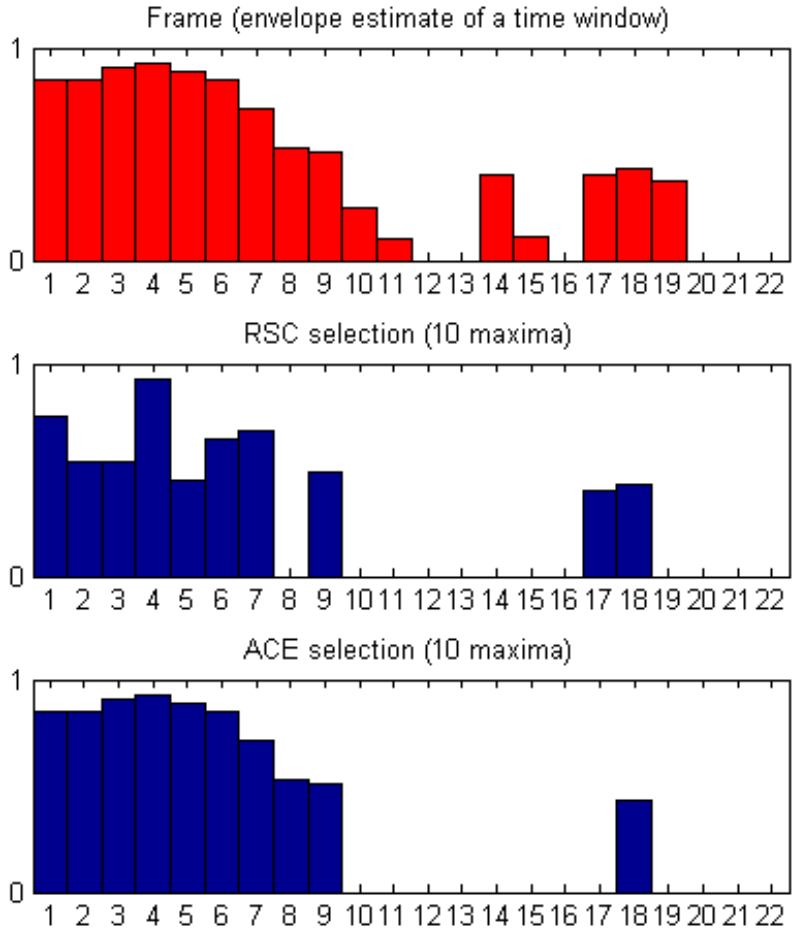


Figure 3.22: *Stimulation selection for ACE and RSC on the same frame, 10 maxima*

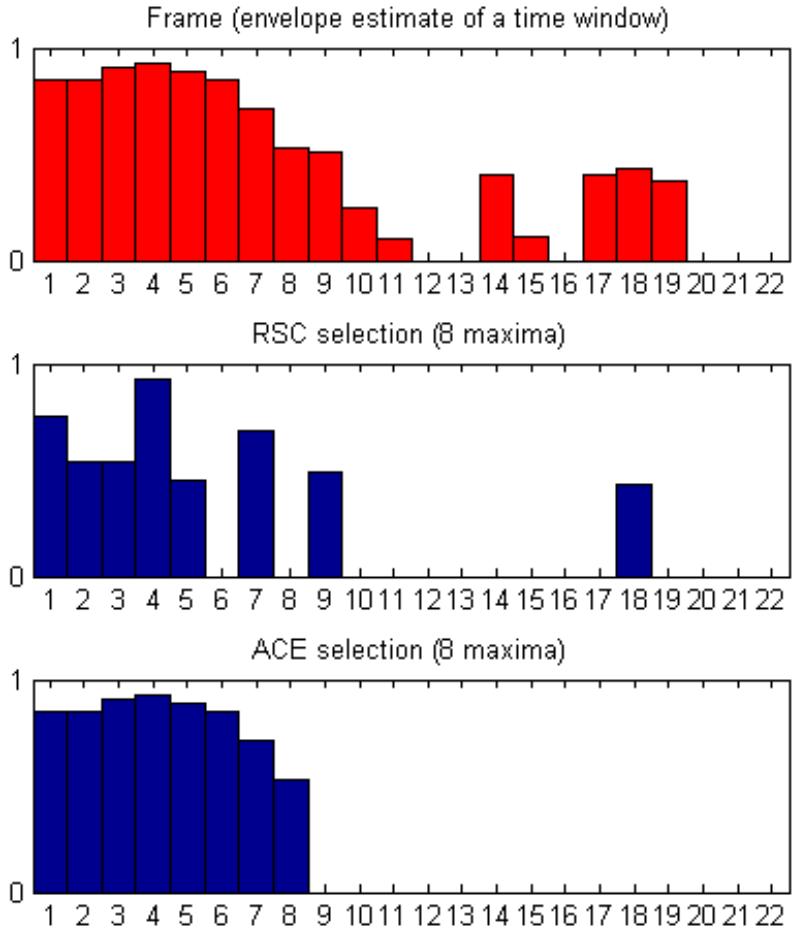


Figure 3.23: *Stimulation selection for ACE and RSC on the same frame, 8 maxima*

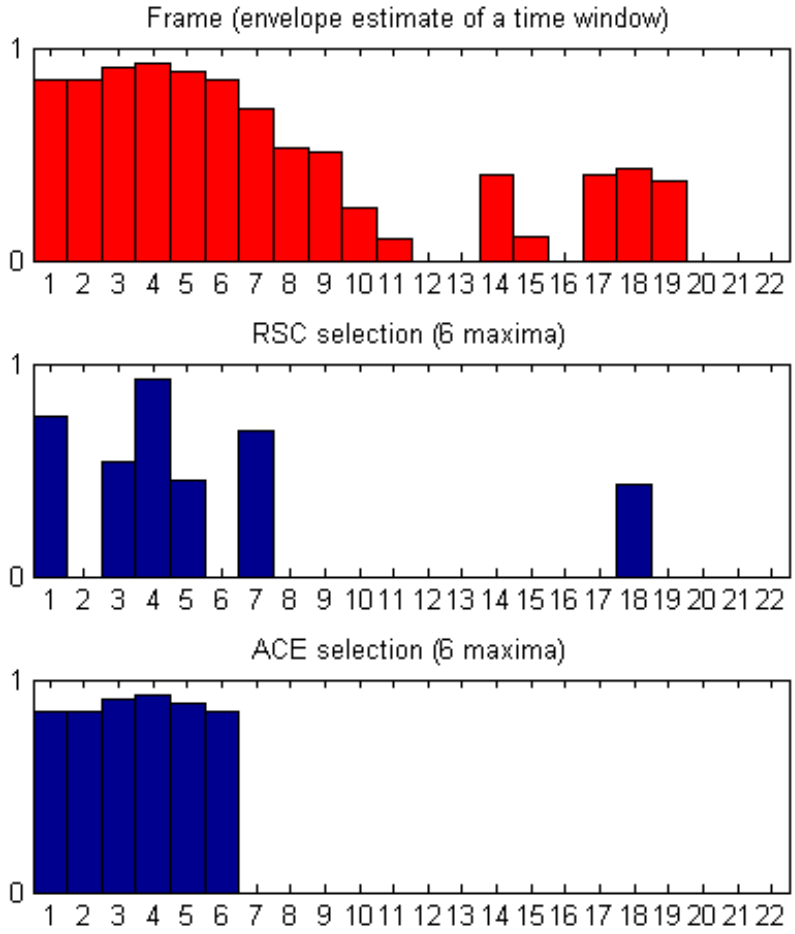


Figure 3.24: *Stimulation selection for ACE and RSC on the same frame, 6 maxima*

3.3.4 Freedom Microphone Calibration

In order to match the NMT output with that of the Freedom SP, we need to perform a calibration. Because we have no direct access to the internals of the Freedom SP, we will indirectly work with the RF output instead. Firstly, we obtain outputs from the Freedom SP with known sinusoidal inputs at calibrated levels. The measurements were done within the sound-proof box of the Aurical audiometer, and the pure tones from 40 to 70 dB SPL-A in 5dB steps were generated using the audiometer. The Freedom SPs RF output was captured using a PCI using the RFcap application. A 25cm transmission cable was used with the Freedom SP in order to enable its RF coil to be led out of the sound-proof box to the PCIs coil receptacle. Note that the Freedom SP had to be programmed using Custom Sound 3.0 in order to be able to use its output with the PCI. Firmware updates with Custom Sound 3.1 have added additional implant ID features which prevent the Freedom SP from sending any output when the correct implant (obviuosly not the implant chip in the PCI) is not detected. The Test Freedom patients Map 136 without any Custom Sound features (eg ASC, ADRO) was used, with the Freedom SP set to the default sensitivity of 12. Note that it is assumed that the Freedom SPs microphone is in good working order and matches specifications. The captured outputs are then analyzed to find the channel with the largest Current Level, and the modal CL value from this electrode is then taken to be the corresponding output (see Figure 3.25).

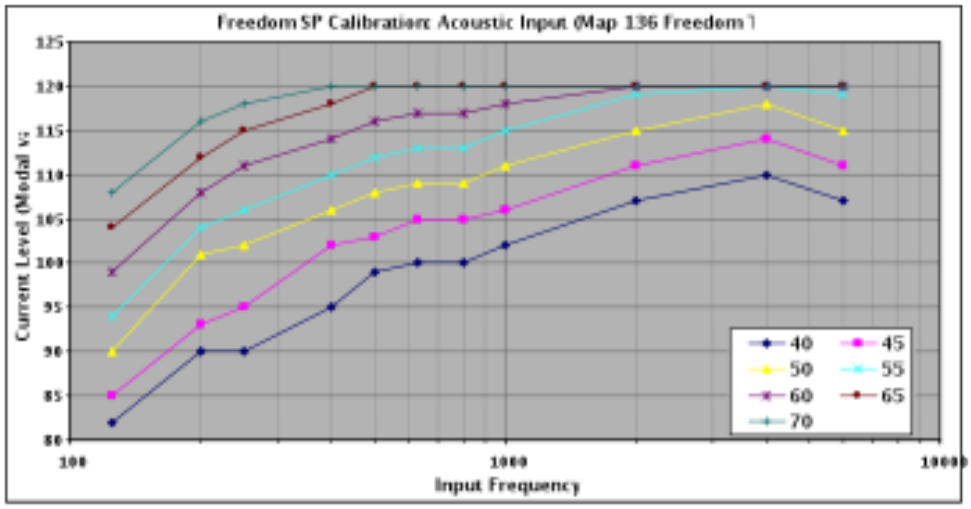


Figure 3.25: *Frequency Response Estimate of Freedom Microphone, obtained by taking the modal value of the implant channel with the highest current level. Legend displays the dBSPL level of the played sine wave.*

Next, full-scale sinusoidal WAV files were generated for the corresponding frequencies

(125, 200, 250, 400, 500, 630, 800, 1000, 2000, 4000 and 6000Hz) and sent to the default ACE speech coding strategy, using the same Test Freedom map 136. The Gain parameter was used to attenuate the input WAV file in 5dB steps. The resultant output was then analyzed in the same manner to produce NMTs corresponding output in Current Levels. Note that the `map.source` parameter had to be set to Headset in order to invoke the Sprint Front-End functions which corrects for the frequency response of the Sprint SPs HS8 microphone and additional front-end circuitry.

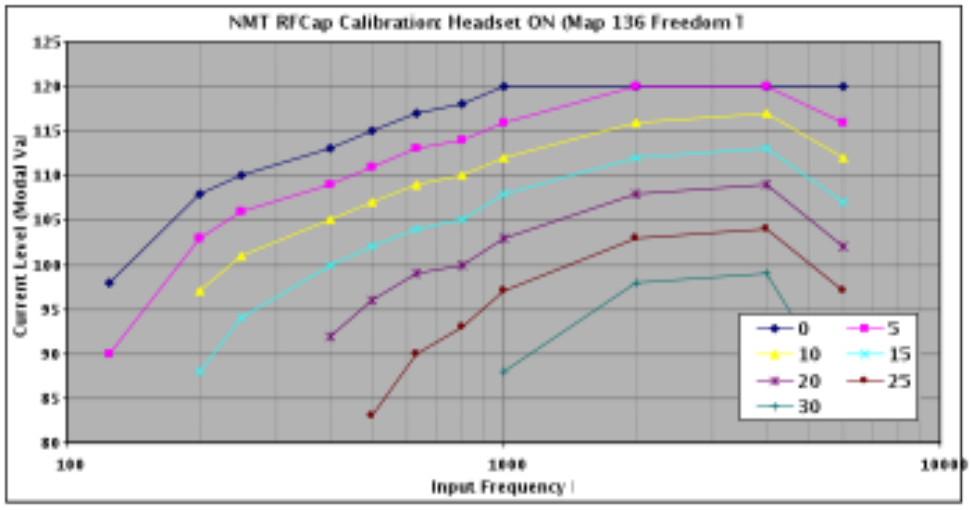


Figure 3.26: *Frequency Response Estimate of Freedom Microphone, with HS8 Microphone compensations. Legend displays the attenuation levels applied in dB.*

The output with the SPrint emulations does not quite match the output of the Freedom SP (see Figure 3.27)

We observed that the frequency response estimates from the two methods are impossible to properly match with simple linear gain. The function shapes are different, so a calibration filter had to be constructed in order to compensate for this discrepancy. For each frequency, the gain was adjusted so that the output (with the full-scale sinusoidal input WAV file) corresponds to the calibration output levels from 70dB SPL-A downwards. The unattenuated full-scale sinusoidal WAV file was therefore assumed to correspond to an input level of 70dB SPL-A. The corresponding gain values were then noted.

These corresponding gain values therefore specify the filter required to match the unattenuated full-scale sinusoidal input to the calibrated 70dB SPL-A outputs. The calibration filter was implemented using the Matlab `fir2()` function to produce an 80^{th} order FIR filter with the corresponding frequency response, which can be viewed as an interpolation between the data points measurements were made at. Furthermore, the 14dB (factor of

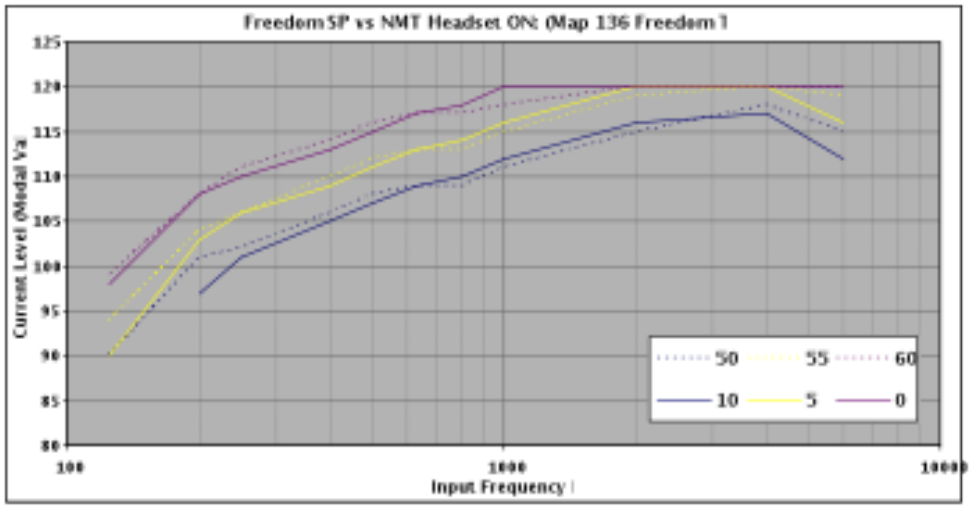
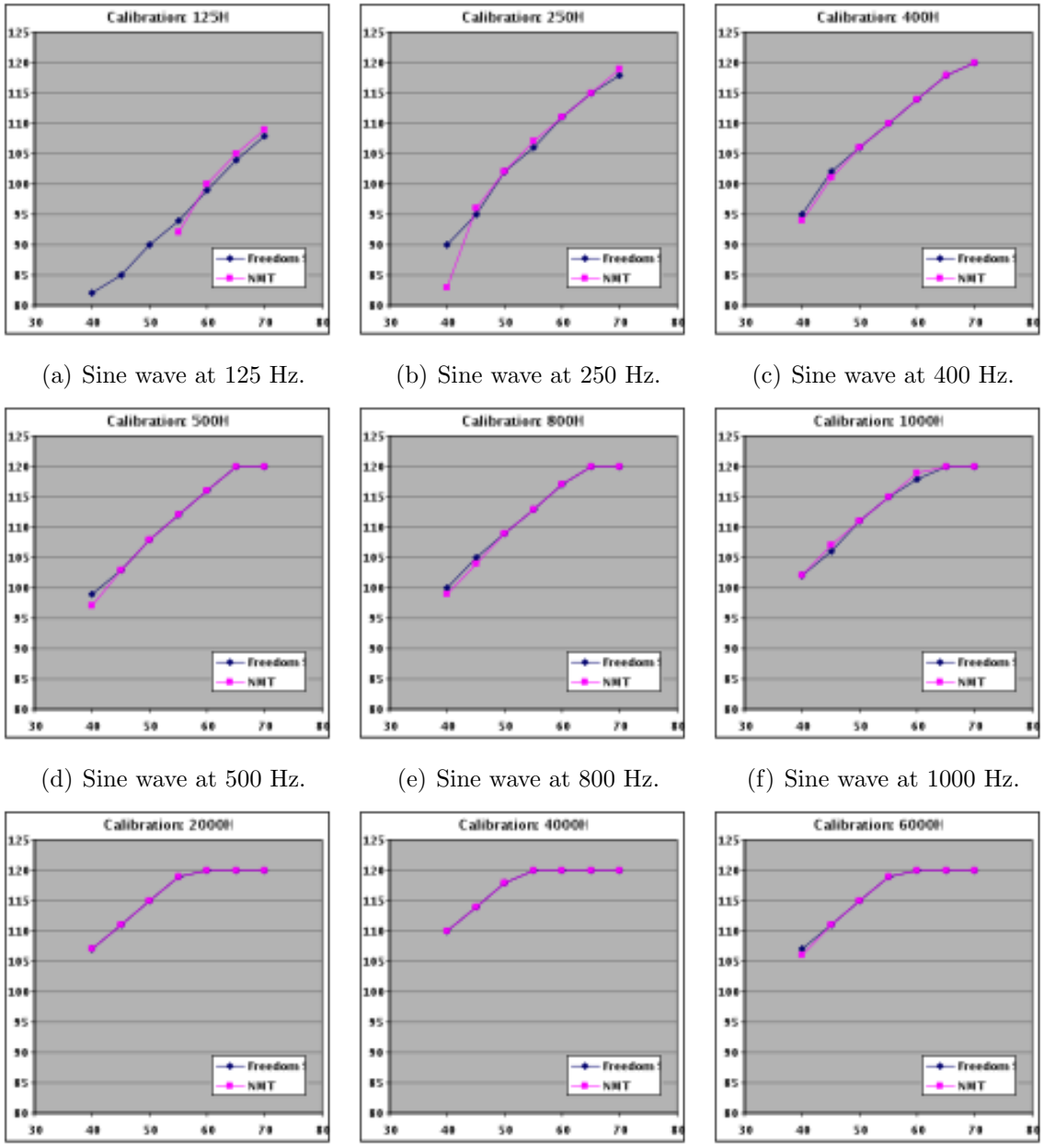


Figure 3.27: *Comparison of HS8 and Freedom microphone characteristics. Dotted lines represent the sound level of played sine waves in dBSPL, straight lines represent the attenuation levels applied to full-scale sine waves in dB.*

25.12) required for the filters maximum value is implemented in the Matlab code below, and the filter frequency response is displayed in Figure 3.29. The calibration filter has positive gain for all the spectrum, but boosts the lower frequencies slightly more and high frequencies significantly more, in order to compensate for the measured discrepancies.

```
f = [0 125 250 500 800 1000 2000 4000 6000 8000];
f = f/8000; % assumes fs = 16000
hcalib = [11 11 11 11 9 9 9 12 14 14];
% Calibration values for ECC/RSC at 70dB SPLA
m = hcalib/max(hcalib);
B = fir2(80,f,m); % 80th order arbitrary shaped filter
B = B * 25.1189;
% fir2 output is normalized to max 1dB, we need max 14dB
A = 1;
% filter parameters are A (denominator) and B (numerator)
% filter the input signal with the given filter
u = filter(B,A,signal);
% remove phase change by the filter
b = sign(real(u));
%extract real numbers sign (Remove phase information)
b1 = abs(u);       %calculate the absolute of the complex value
signal = b.*b1;    %multiply with the sign
```



(a) Sine wave at 125 Hz. (b) Sine wave at 250 Hz. (c) Sine wave at 400 Hz.

(d) Sine wave at 500 Hz. (e) Sine wave at 800 Hz. (f) Sine wave at 1000 Hz.

(g) Sine wave at 2000 Hz. (h) Sine wave at 4000 Hz. (i) Sine wave at 6000 Hz.

Figure 3.28: *Calibration of sound pressure levels to implant current levels at different frequencies*

3.4 Implementation

The implementation of the RSC algorithm was realized in Matlab. Discrete-time equivalents of the equations described in §3.3.3 are used in the software. It is built to be

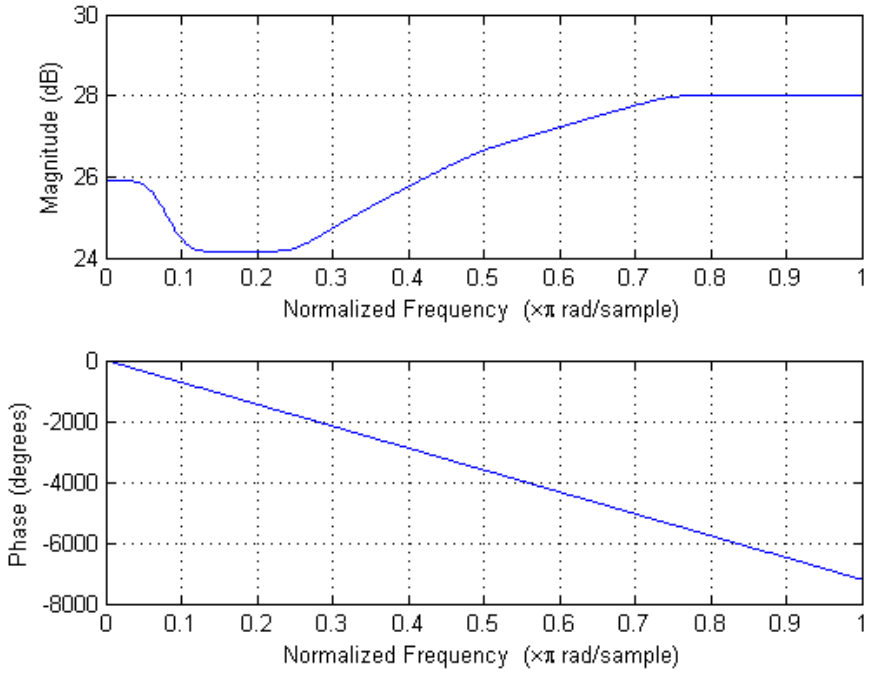


Figure 3.29: *Frequency response of Freedom calibration filter*

compatible with the existing implementation of research ACE strategy available in Nucleus Matlab Toolbox from Cochlear Corporation [4].

3.4.1 RSC Toolbox

Strategy Processing Functions

A block diagram of the main processing component of RSC Toolbox is given in Figure 3.30. Unshaded rectangles represent functions, shaded ellipses represent variables. Main signal path is highlighted in red. This is reflected in the default value of the `processes` field of a parameter struct generated by `RSC_map()`:

```
1 >>p.source = 'Freedom';
2 >>p = RSC_map(p);
3 >>p.processes
4
5 ans =
6
7     'Wav_proc'
```

```
8       'Freedom_microphone_proc'
9       'SPrint_front_end_proc'
10      'Input_scaling_proc'
11      'FFT_filterbank_proc'
12      'Power_sum_envelope_proc'
13      'Gain_proc'
14      'LGF_proc'
15      'RSC_proc'
16      'Channel_mapping_proc'
```

To process a .wav file after having set up the map variable as above, we simply need to run

```
1 >>q = Process(p,wavFilename);
```

For detailed function descriptions, see Appendix.

RSCGUI

We developed a GUI (Graphical User Interface) for easy inspection of how parameters affect the strategy output. Called `rscgui` it lets the user process any .wav file, vary map and model parameters and inspect pulse sequence plots and statistics of pulse sequences. The main window of the GUI is presented in Figure 3.31. We will present `rscgui`'s functionality piece by piece by going over the annotated regions in Figure 3.31:

1. Map selection: `rscgui` connects to the UniversitätsSpital Zürich (University Hospital Zürich) database to get implant maps belonging to specific patients, artificial, theoretical maps are also available. These maps are used as the basis for processing signals.

2. Input controls: Lets user load a .wav file, listen to it and plot it.

3. Edit parameters panel: Displays relevant parameter values from the loaded map. The user can edit them on-the-fly and (re-)process the signal to inspect changes. Three groups of parameters are available for modification: Strategy parameters (number of maxima, channel stimulation rate), pulse parameters (phase width, phase gap) and loudness mapping parameters (Q, base level).

4. Sequence type selection: Determines whether a channel-magnitude sequence (values in $[0, 1]$, represents fraction of channel) or a pulse sequence (values in $[0, 255]$, represents current levels internal to implant) will be displayed.

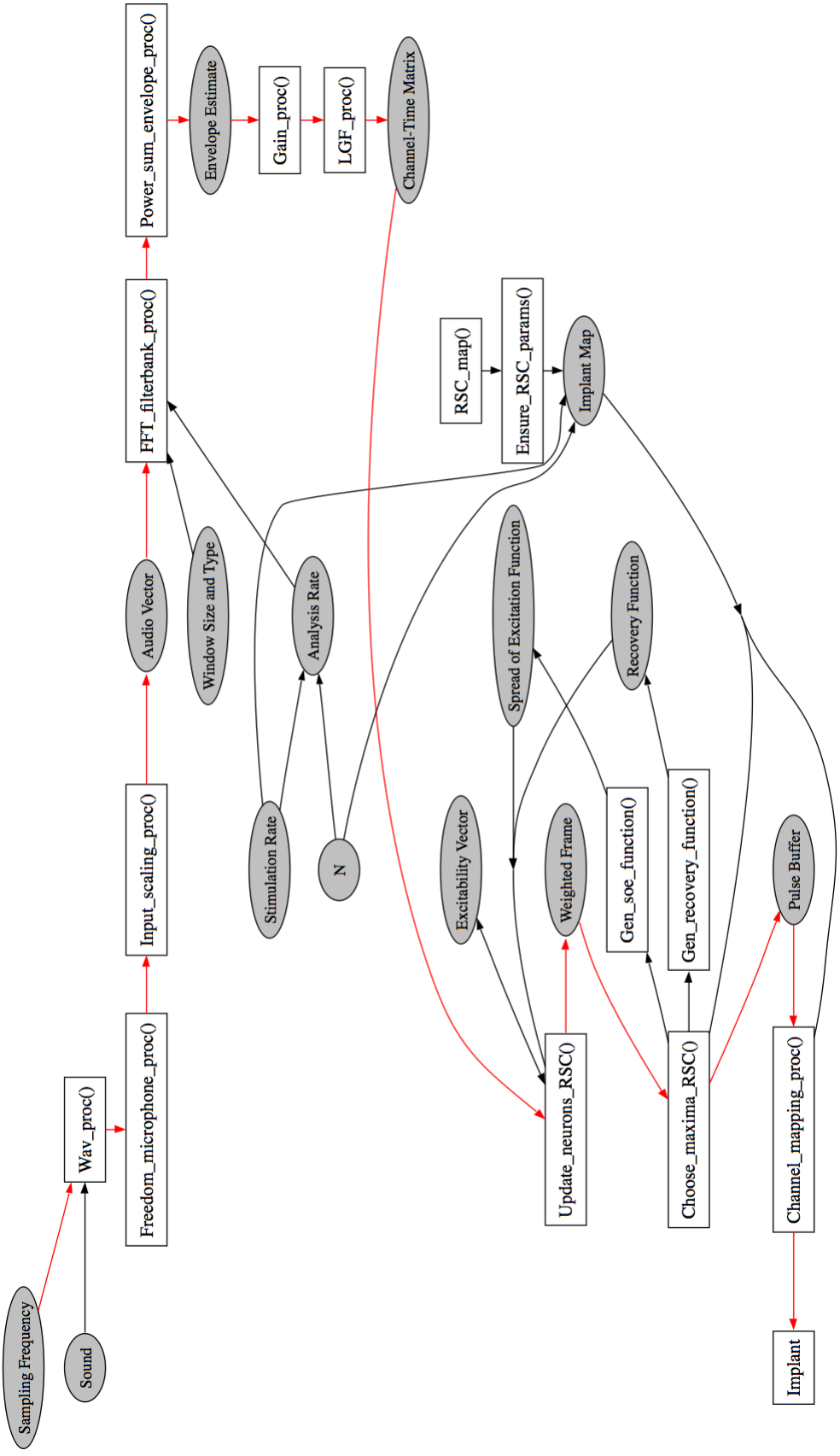


Figure 3.30: *Block diagram of RSC Toolbox. Unshaded rectangles represent functions, shaded ellipses represent variables. Main signal path is highlighted in red.*

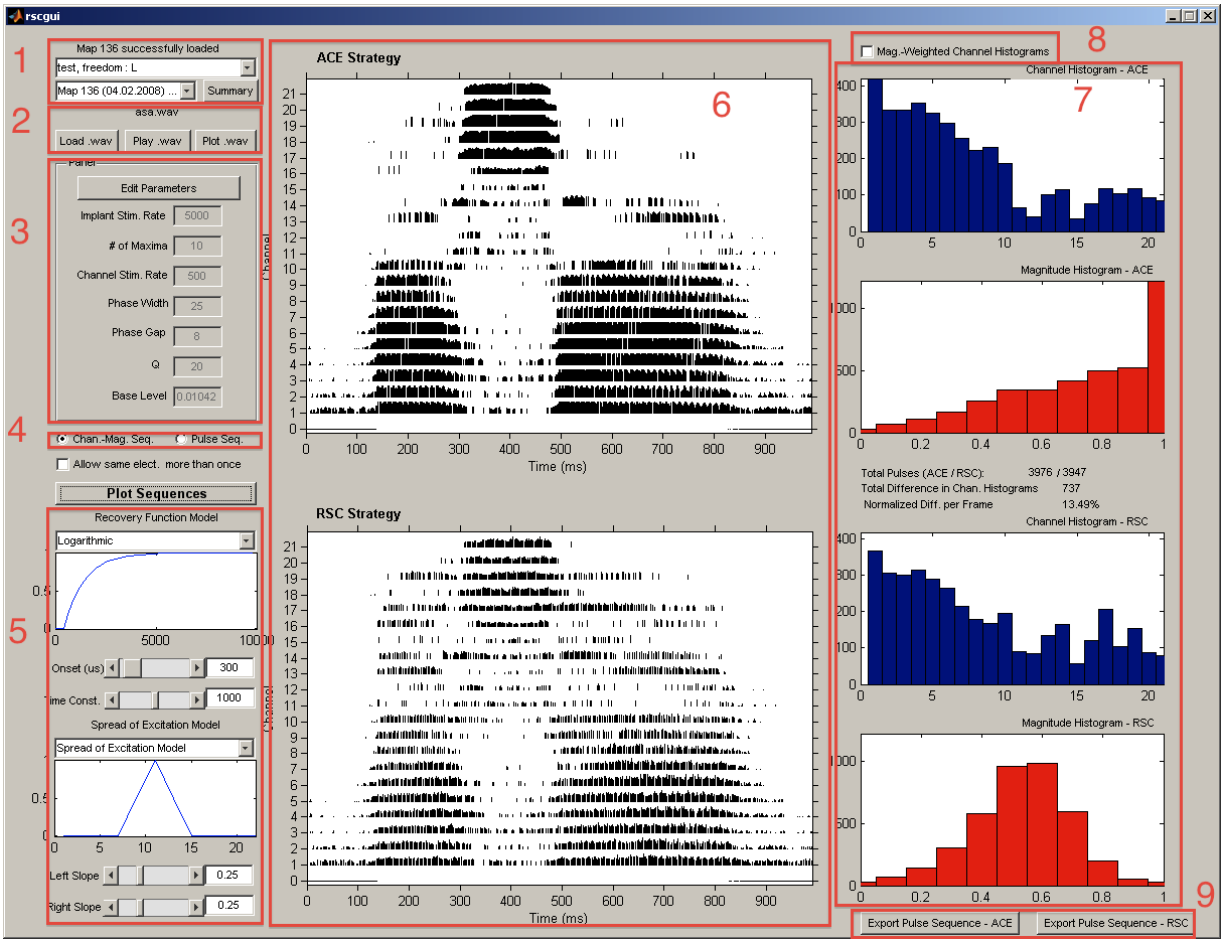


Figure 3.31: *Main window of rscgui*

5. Model parameters: Displays the recovery and spread of excitation functions from the loaded map. User can change the model type for these functions, as well as parameters that define function shape.

6. Sequence displays: Displays the generated channel or pulse sequence. The display is divided into horizontal stripes with equal heights, each representing a channel. Black vertical lines in these stripes represent the channel or pulse magnitude with their height and the temporal window in which they were triggered with their horizontal position. Top axes display the sequence generated with ACE and the bottom one with RSC.

7. Channel and magnitude histograms: Displays two kinds of histograms obtained from the generated sequence. Channel histogram is a histogram of pulse occurences per channel. Magnitude histogram is a histogram of channel or pulse magnitudes with histogram bins of 0.1 and 1, respectively. The top pair is for the ACE sequence and

the bottom for the RSC sequence.

8. Magnitude weighted channel histograms: (See item (7) in this list.) Instead of summing the number of occurences for each channel, sums the channel magnitudes in channel histograms. The resulting "Magnitude-weighted channel histogram" is an estimate of delivered energy per channel.

9. Export sequence: Writes the generated sequence to disk in .xml format. ACE and RSC sequences can be saved separately.

`rscgui` provides the option to process the signal with both ACE and RSC strategies or only RSC. It displays both pulse sequences at the same time, facilitating comparisons with an ACE baseline.

EXPGUI

As explained in §4.1.2, we required custom software to present stimuli to test subjects. `expgui` was written to fill this need. It contains the same database map module as `rscgui`. ACE and RSC options along with number of maxima options are available. An important feature of `expgui` is the map modification module. It lets the user make global modifications to T-, C- or T- and C- levels, as a way of controlling loudness. Any signal can be processed with this modified map and presented to the test subject, allowing for tuning of loudness by feedback from the subject. Once a satisfactory map is reached, `expgui` can then process tokens for the desired test (C12 or Insrec8 are currently available), keep them in memory and present them to the test subject as per the desired number of repetitions and a randomized order. The results are saved in a plain text (.txt) file for potential compatibility with other software. In addition, all pulse sequences used in the test, as well as the modified maps are saved in categorized folders for future reference.

Chapter 4

Experiment

4.1 Study Design

4.1.1 Hypothesis

We hypothesize that by modeling the refractory behavior, RSC is able to deliver more natural sound than ACE. The assumption underlying this hypothesis is that by not stimulating the sites that are not able to fully receive the stimulus, we avoid unnecessary overstimulations. This results in two desired effects: The sites in refractory period are allowed to repolarize, being better able to receive stimulus in the short-term future; and that by stimulating other sites instead of overstimulation of the same site, the available spectrum is better utilized.

Before starting experimentation, we inspected and compared various sequences generated using ACE and RSC strategies. We observed that RSC is indeed better at utilizing the available spectrum as the number of maxima per frame is decreased. Another observation we made is that current levels, and consequently energy expenditure is much lower with RSC. If the performance of RSC would prove similar to ACE, the power savings would be a clear advantage. These observations are illustrated in Figures 4.1- 4.4

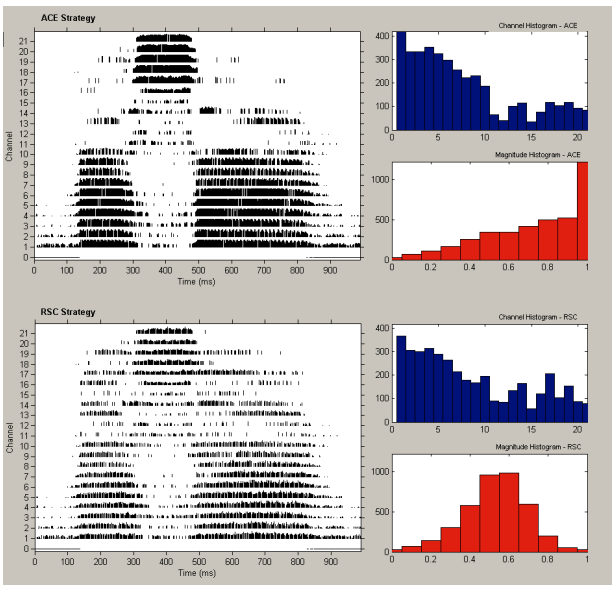


Figure 4.1: *Comparison of channel-magnitude sequences with corresponding channel and magnitude histograms, for 'asa' token, 10 maxima*

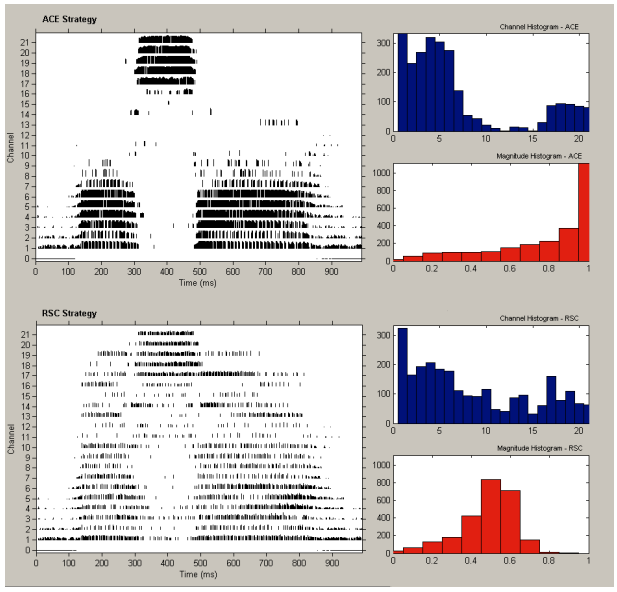


Figure 4.2: *Comparison of channel-magnitude sequences with corresponding channel and magnitude histograms, for 'asa' token, 6 maxima*

It is worth noting that even at 2 maxima, the spectrum of the signal is considerably

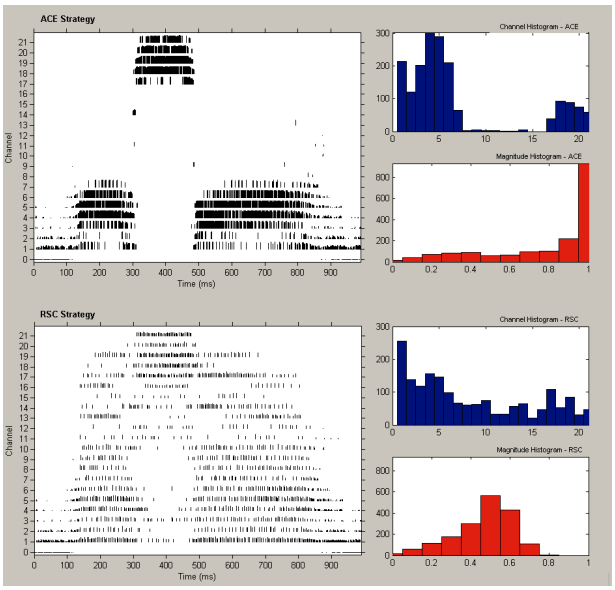


Figure 4.3: *Comparison of channel-magnitude sequences with corresponding channel and magnitude histograms, for 'asa' token, 4 maxima*

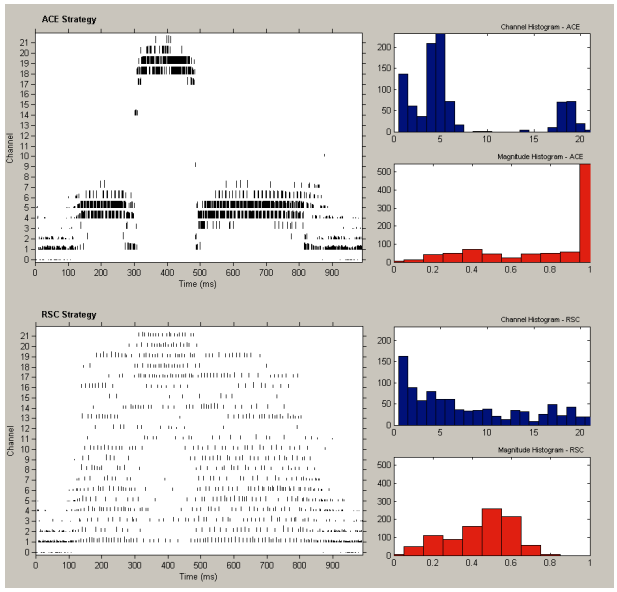


Figure 4.4: *Comparison of channel-magnitude sequences with corresponding channel and magnitude histograms, for 'asa' token, 2 maxima*

well-represented by RSC.

4.1.2 Experimental Protocol

Although we hypothesize to obtain "more natural sound", which is a qualitative assessment, we opted to carry out recognition tests. The motivation for this is the difficulty of obtaining and evaluating statistically significant qualitative data, and the assumption that "more naturalness" will lead to increase in recognition performance.

To test our hypothesis, we initially proposed a C12 Consonant aCa-Logatome (C12 for short) test and a closed-set recognition test of 8 musical instruments playing the same phrase (InsRec8 for short).

In the pilot trials, it became apparent that the InsRec8 test is too difficult for untrained subjects, leading to obstructions in making a clear conclusion. We decided to keep the small number of InsRec8 tests we did as a pilot trial, and removed InsRec8 from the experimental protocol in the interest of better utilizing limited test-subject time.

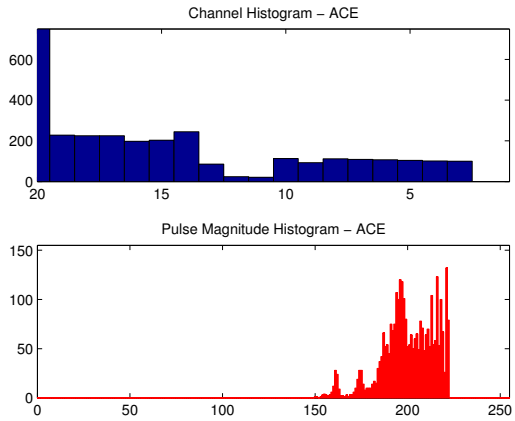
In our initial protocol, we proposed three conditions for C12: ACE with 10 selected maxima (ACE10), RSC with 10 selected maxima (RSC10) and RSC with 4 selected maxima (RSC4). ACE10 would be the baseline, against which to compare the RSC performance.

We proposed that RSC4 would yield comparable performance to ACE10, and RSC10 would yield equal or better performance. However, after pilot trials, we observed that even RSC10 performed worse than ACE10, and we removed RSC4 from further tests.

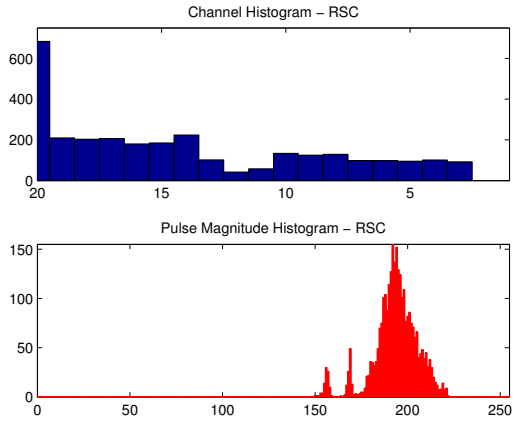
We tested 9 native German-speaking adults with the C12 test. In C12, the subject is played back one of the 12 aCa tokens and chooses the answer from a touch screen, on which all tokens are displayed as buttons. We presented 4 repetitions of each token in random order, for a total of 48 presentations per condition.

One of the other considerations in the experimental protocol was the choice of software. MACarena [8] [5] is an existing software solution that is commonly used in this line of experimenting. We opted to use a custom-written test software instead of MACarena. The reason for this is that in RSC-coded stimuli, we observed a decrease in loudness that was difficult to reliably predict. This required us to make on-the-fly adjustments to loudness. MACarena requires that the stimuli be in .xml format, and even using optimized .xml exporting functions from Nucleus Matlab Toolbox, the export duration was much longer than the available test subject time. Thus, our solution was to write test software that mimics MACarena in test interface and direct-to-implant streaming, but keeps the stimuli in computer memory, so as to make quick modifications. We did not encounter loudness problems with ACE tests, since they are calibrated using the Freedom calibration filter (see §3.3.4). For RSC tests, loudness modifications were done by making global additions or subtractions to threshold and comfort levels.

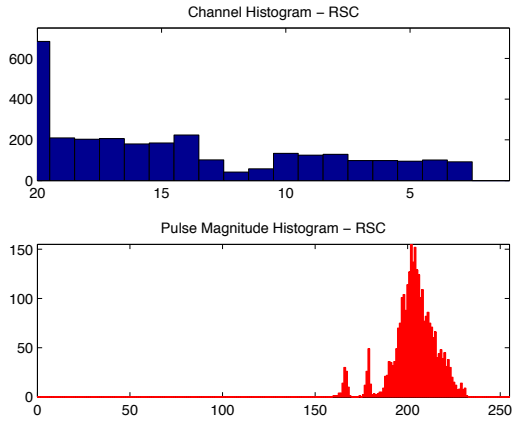
In making global loudness changes, the question arises, does the maxima selection change due to changes in loudness? Since we only offset the existing levels by the same amount, and since the pulse magnitude remapping (3.14) is a linear operation, the channel magnitudes are independent from T- and C-levels. Thus the loudness modifications do not affect maxima selection (see Figure 4.1.2).



(a) ACE10, no map modification



(b) RSC10, no map modification



(c) RSC10, T- and C-levels increased

Figure 4.5: *Histograms of occurences of pulses in channels (channel histogram) and occurences of pulse magnitudes from a pulse sequence generated from the C12 'asa' token; with ACE10 and RSC10, in (a) and (b),(c), respectively. Notice how an increase in T- and C-levels only shifts the pulse magnitude right and the shape stays the same.*

Chapter 5

Results

5.1 Results

The results are evaluated in terms of recognition performance. As mentioned before we only present Insrec8 results as a pilot study and do not draw any conclusions from them. Summary of results are presented in tables and figures below. We also present more detailed "confusion matrices", see Appendix.

Table 5.1: *Correct recognition rates of C12 Tests*

	ACE10	RSC10	RSC4
GZ	83.33%	45.83%	52.08%
TB	58.33%	39.58%	-
JK	64.58%	52.08 %	62.50%
KS	66.67%	54.17%	-
AB	83.67%	85.42%	58.33%
WH	77.08%	83.33%	75.00%
KW	64.58%	54.17%	47.92%
CE	37.50%	47.92%	41.67%
MB	47.92%	-	41.67%
Mean	64.85%	51.39%	42.13%
Std. Dev.	15.51%	25.04%	26.07%

Table 5.2: *Correct recognition rates of InsRec8 Tests*

	ACE10	RSC10	RSC4
GZ	75.00%	-	-
TB	31.25%	31.25%	-
JK	-	-	-
KS	46.87%	31.25%	-
AB	78.12%	-	40.62%
WH	-	-	-
KW	59.37%	21.87%	37.5%
CE	-	-	-
MB	-	-	-
Mean	64.85%	57.81%	54.17%
Std. Dev.	15.51%	17.09%	12.08%

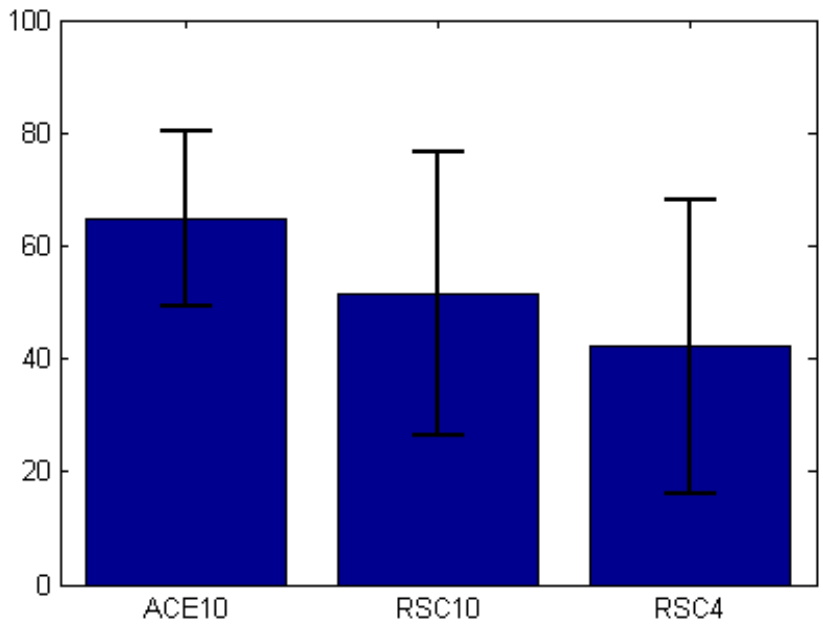


Figure 5.1: *Correct recognition percentage means per C12 test condition (error bars indicate standard deviation)*

5.2 Discussion

The results show that RSC in its current state underperforms ACE, contrary to our initial assumptions. Under this light, we attempt to explain if and where our assumptions were wrong and why ACE performs better.

As discussed in §4.1.2, using the same implant map parameters, RSC is significantly less loud than ACE. Test subjects also described the sound as "muffled". We believe this is because of the under-representation of the signal due to spread of excitation and recovery functions. Using functions that will lead to less severe reductions in the stimuli may improve this slightly. However we do not expect large improvements, because similar results are reported using the MP3000 strategy [9], which reduces the signal representation less severely than RSC, since it only uses the spread of excitation function and no recovery functions.

ACE strategy presents denser stimulation than RSC. From the RSC perspective, the signal is over-represented in certain stimulation sites and under-represented in others. This over-representation might actually be desirable. After a strong stimulus, a larger fraction of neurons closer the center of the electric field are in depolarized states, while

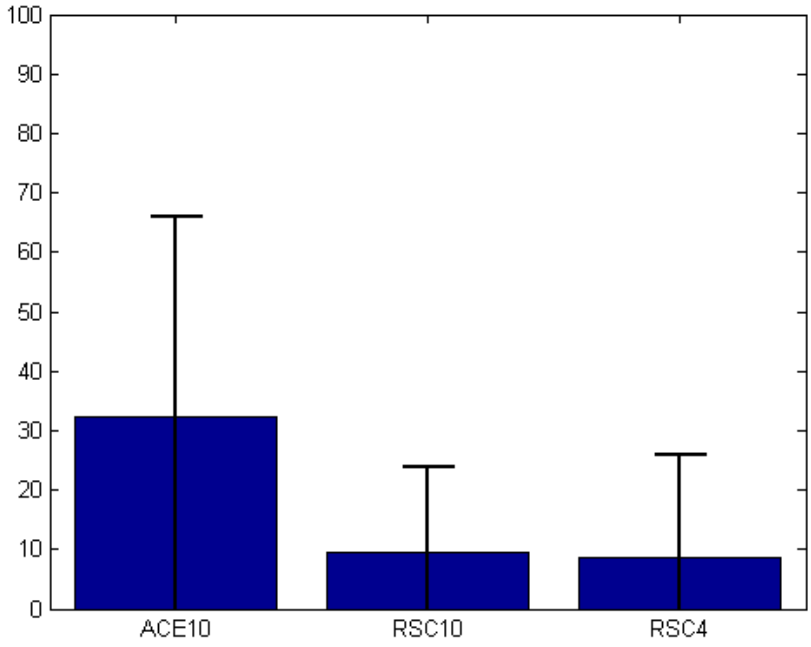


Figure 5.2: *Correct recognition percentage means per InsRec8 test condition (error bars indicate standard deviation)*

a smaller fraction of neurons closer to the periphery of the electric field display the same behavior. As such a number of these neurons closer to the periphery are still available for stimulation. In RSC, we disregard these neurons due to our assumption that most of the site is in absolute refractory period; whereas in ACE, by "over-representing" the signal, these neurons are utilized for sound transmission in consequent frames. If indeed the contribution of these neurons are more significant than we initially proposed, it would affect recognition performance, as well as help explain the decrease of loudness in RSC compared to ACE.

When performing tests, we compensated for the decreased loudness by increasing threshold and comfort levels of the implant map. This is a method of compensation also used in previous MP3000 studies **??**. This increase in levels unavoidably leads to spread of excitation functions with larger amplitudes, and may increase undesired channel interaction, consequently decreasing performance. Although the RSC model attempts to address this problem, due to the linearity of the used operations, an offset to both T- and C-levels does not change the maxima selection, and fails to address increased channel interaction. It may be a worthwhile line of inquiry to re-interpret what T- and C- levels should mean in the context of RSC.

Taking into consideration these issues about loudness, a denser stimulation pattern from the coding strategy seems to be more desirable. However the optimal reduction from ACE or the increase from RSC or MP3000 remains unknown at present.

In our tests, we used the same SoE and recovery functions based on theoretical models for all subjects. We did not investigate the usage of individually-measured SoE and recovery functions thoroughly, although we did perform pilot trials. We did not observe an improvement with the individual functions over the theoretical ones. We do not expect significant improvements with finely customized functions, but this could be a future investigation.

Also of note is the fact that our SoE functions are not level-dependent. Even though we apply linear scaling to SoE functions based on the level, we do not account for changes in function shape that might be happening on the physical level. This may be a contributing factor to performance.

Another factor we must consider is the training effect. Most of our subjects preferred and used ACE in daily life. As such, they are expected to perform better using ACE.

Chapter 6

Conclusion and Future Work

In its current state, RSC performs slightly worse than ACE in terms of logatome recognition. The MP3000 user we tested did perform slightly better with RSC. Future tests with separate ACE/MP3000 groups could show if such a trend exists.

RSC has been described as "muffled" by ACE users, lowering the chances that it will perform better with music for ACE users.

We've also observed, through ACE users performing better and the loudness decrease in ACE, that RSC assumptions need to be re-evaluated. Basic psychophysical experiments are needed to check whether the "over-stimulation" (stimulus within the refractory period window) of the same site by ACE has perceptual effects. Depending on this, the RSC assumption about redundancy may prove obsolete. Such experiments could also help determine the exact redundancy. This would help optimize the redundancy removal in RSC, leading to louder and less muffled stimuli for ACE users.

Acknowledgements

This study has been made possible by UniversitätsSpital Zürich, ORL Department, Laboratory of Experimental Audiology. The author would like to thank Dr. Wai-Kong Lai for his continued support and guidance and Dr. Steven Schimmel for suggesting the name "Refractory State Coding".

Appendix A

Confusion Matrices

We visualize the C12 test results in the form of confusion matrices. The vertical axis represents the presented token (question) and the horizontal axis represents the answer. The figures are divided into colored regions by number of occurences which are also displayed in the regions. For all answers to be true, the diagonal of the figure should have the value 4 (since that was the number of repetitions used in the test). Wrong answers can be observed as deviations from the diagonal. Each tick in the axes is marked with the consonant of the C12 token.

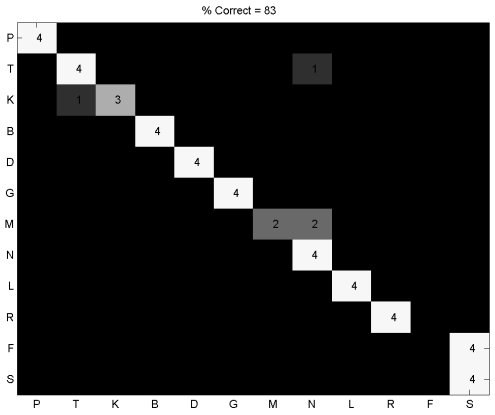


Figure A.1: *Confusion Matrix for AB, ACE10*

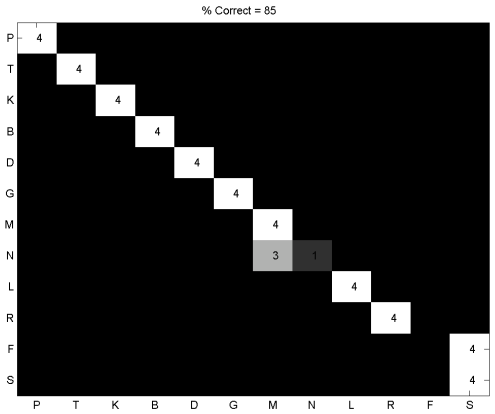


Figure A.2: *Confusion Matrix for AB, RSC10*

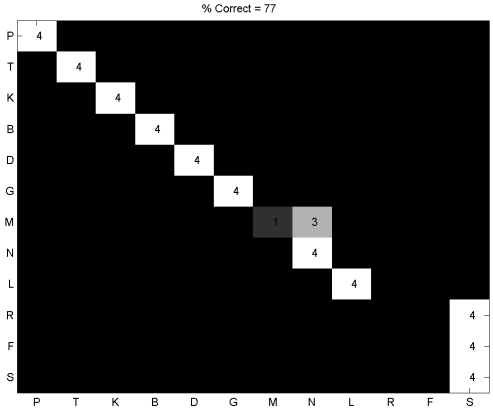


Figure A.3: *Confusion Matrix for WH, ACE10*

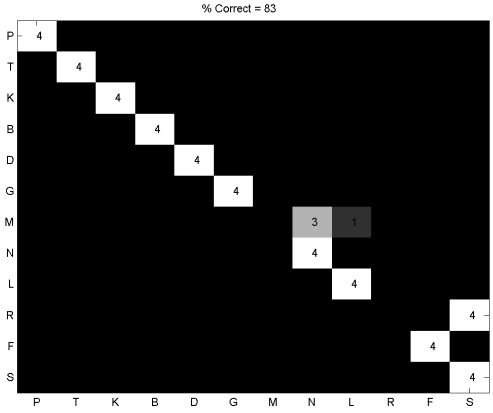


Figure A.4: *Confusion Matrix for WH, RSC10*

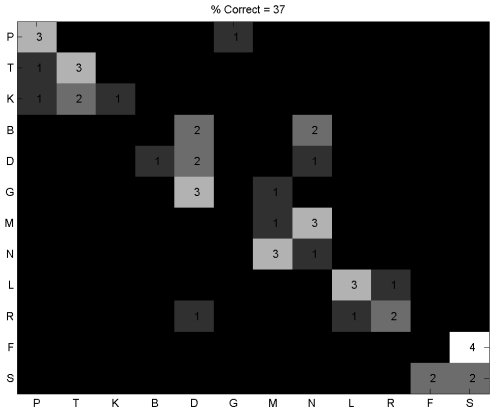


Figure A.5: *Confusion Matrix for KW, ACE10*

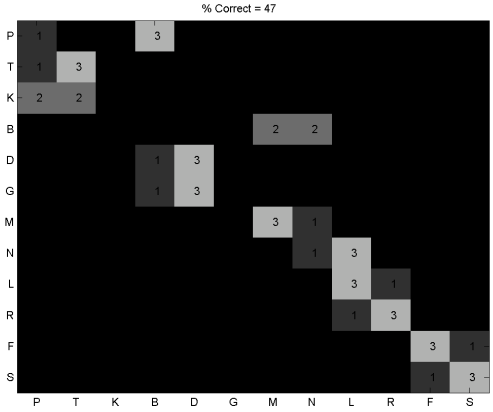


Figure A.6: *Confusion Matrix for KW, RSC10*

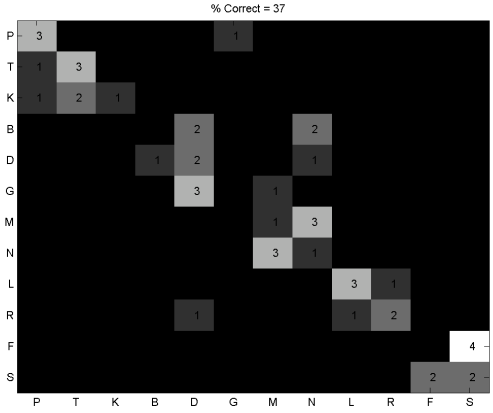


Figure A.7: *Confusion Matrix for CE, ACE10*

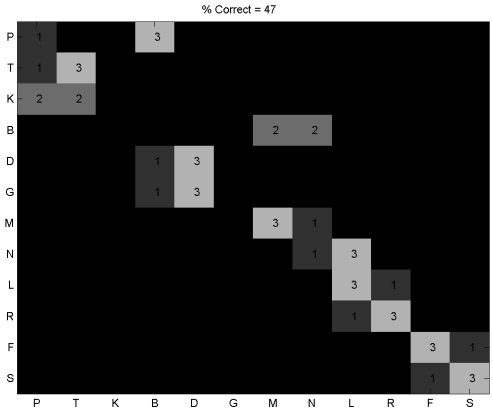


Figure A.8: *Confusion Matrix for CE, RSC10*

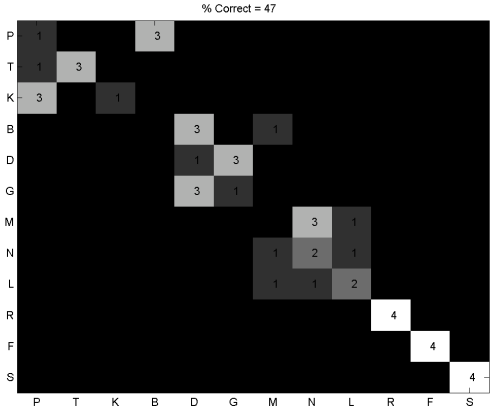


Figure A.9: *Confusion Matrix for MB, ACE10*

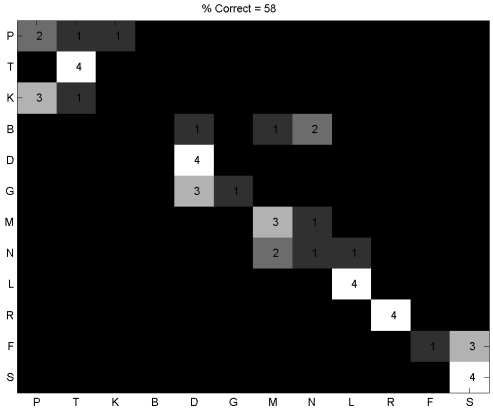


Figure A.10: *Confusion Matrix for TB, ACE10*

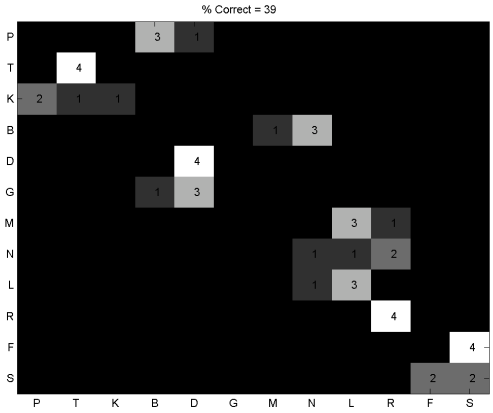


Figure A.11: *Confusion Matrix for TB, RSC10*

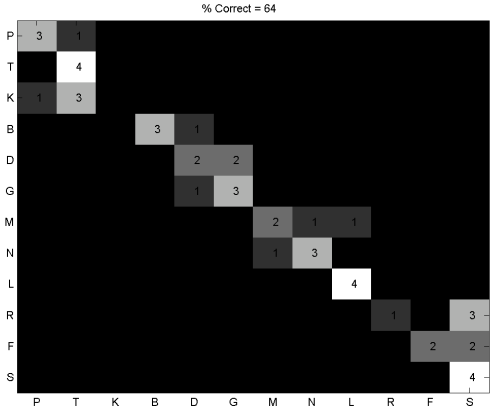


Figure A.12: *Confusion Matrix for KJ, ACE10*

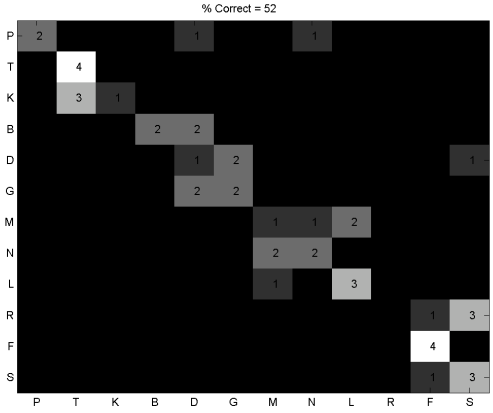


Figure A.13: *Confusion Matrix for KJ, RSC10*

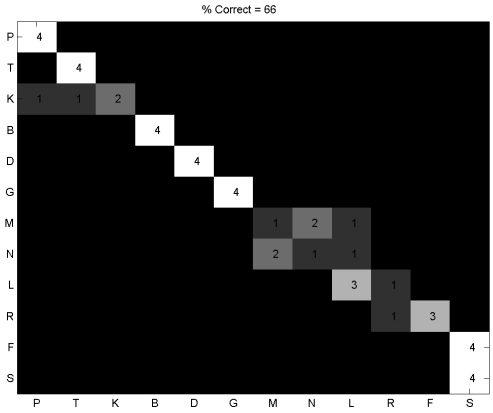


Figure A.14: *Confusion Matrix for KS, ACE10*

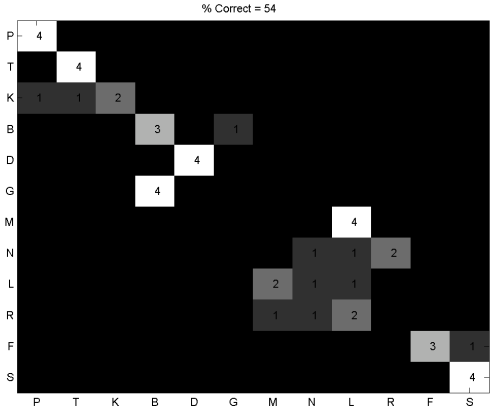


Figure A.15: *Confusion Matrix for KS, RSC10*

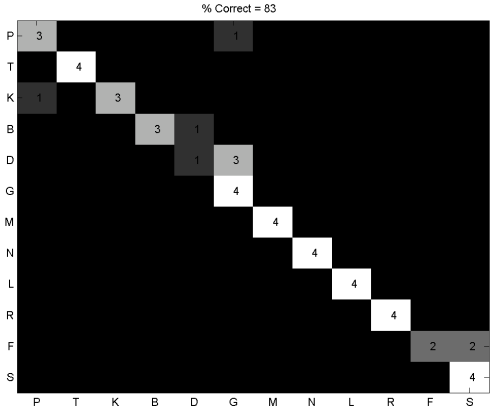


Figure A.16: *Confusion Matrix for GZ, ACE10*

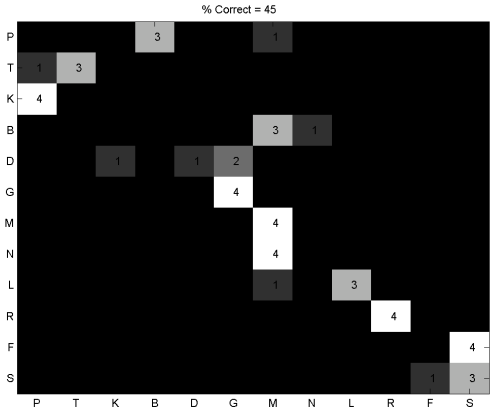


Figure A.17: *Confusion Matrix for GZ, RSC10*

List of Tables

List of Figures

Bibliography

[1] P Abbas, M Hughes, and C Brown. Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential. *Audiology and Neuro-Otology*, Jan 2004.

[2] B Gantz CJ Brown, PJ Abbas. Electrically evoked whole-nerve action potentials: data from human cochlear implant users. *J Acoust Soc Am*, pages 88–1385–91, 1990.

[3] Cochlear Corporation, Lane Cove, New South Wales Australia. Ace and cis dsp strategies. *Software Requirements Specification N95287F*, Issue 1, Oct 2002.

[4] Cochlear Corporation. Nucleus MATLAB Toolbox 2.11. *Software User Manual N95246F*, Issue 1, Oct 2002.

[5] UZH ORL Department. http://www.uzh.ch/orl/projects/speechtests/speechtests.html.

[6] N Dillier, WK Lai, B Almqvist, C Frohne, J Muller-Deile, M Stecker, and E von Wallenberg. Measurement of the electrically evoked compound action potential via a neural response telemetry system. *Annals of Otology Rhinology and Laryngology*, 111(5; PART 1):407–414, 2002.

[7] Q Fu, RV Shannon, and X Wang. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *J Acoust Soc Am*, 104:1–11, 1998.

[8] WK Lai and N Dillier. MACarena: A flexible computer-based speech testing environment. *7th International Cochlear Implant Conference, Manchester, England*, 2002.

[9] WK Lai and N Dillier. Investigating the MP3000 coding strategy for music perception. *11. Jahrestagung der Deutschen Gesellschaft für Audologie, Kiel, 5-8 März*, ISBN 3-9809869-7-7: 1-4, 2008.

[10] P Loizou. Introduction to cochlear implants. *IEEE Engineering in Medicine and Biology*, Jan 1999.

[11] A Morsnowski, B Charasse, L Collet, M Killian, and J Müller-Deile. Measuring the refractoriness of the electrically stimulated auditory nerve. *Audiology and Neurotology*, 11(6):389–402, 2006.

[12] W Nogueira, A Büchner, and T Lenarz. A psychoacoustic nofm-type speech coding strategy for cochlear implants. *EURASIP Journal on Applied Signal Processing*, Jan 2005.

[13] Loizou PC. Mimicking the human ear. *IEEE Signal Processing Mag.*, 15:101–130, 1998.

[14] MW White, MM Merzenich, and JN Gardi. Multichannel cochlear implants. *Arch Otolaryngol*, 110:493–501, 1984.

[15] B Wilson. Signal processing, in cochlear implants. *Audiological Foundations (R. Tyler, ed.),Singular Publishing Group,Inc*, pages 35–86, 1993.

[16] F Zeng, S Rebscher, and W Harrison. Cochlear implants: System design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, 1, Jan 2008.